

연구 · 학술정보 효율적 검색을 위한 온톨로지 기반의 주제 색인어 구조화 방안 연구

A Study on Ontology-based Keywords Structuring for Efficient Information Retrieval

송인석*

In-Seok Song

차 례

- | | |
|---------------------|-----------------------------|
| 1. 서론 | 4. 온톨로지 기반 주제 색인어
구조화 방안 |
| 2. 연구자의 정보활동 및 정보수요 | 5. 결 론 |
| 3. 시소러스 관련 연구동향 | • 참고문헌 |

초 록

본 연구에서는 정보검색도구 관점에서 지식조직체계로서 기존 시소러스 구축방안의 특성과 한계점을 검토하고, 대상 정보의 지식구조의 반영 및 정보 간의 의미관계 추론을 지원하는 온톨로지 기반의 주제 색인어 구조화 방안을 제시한다. 기존의 용어 중심의 시소러스와 달리, 단계별 연구프로세스 과정에서 수행되는 연구자의 정보행위 및 수요 분석에 따라 주제색인어의 개념을 식별·범주화하고, 인문사회과학 분야 학술논문의 지식체계를 구성하는 그 개념들 간의 유기적 관계정의를 통해 주제 색인어 집합의 의미구조를 정형화하였다. 이를 기반으로 각각의 온톨로지 기반 주제 색인어 집합은 구조화된 의미 색인으로서 대상 문서의 지식체계를 표현한다. 정보수요에 따라 정의된 공리나 추론규칙을 활용하여 이용자는 문제 해결에 적합한 정보를 대상 정보의 의미관계로 구성된 주제 도메인의 학술커뮤니케이션 네트워크상에서 분석적 정보탐색을 통해 효율적으로 검색 할 수 있다.

키 워 드

지식조직체계, 시소러스, 온톨로지, 의미관계, 추론, 의미 색인, 분석적 정보탐색

* 한국과학기술정보연구원 선임연구원

(Senior Researcher, Korea Institute of Science and Technology Information, sis@kisti.re.kr)

· 논문접수일자 : 2008년 11월 11일

· 게재확정일자 : 2008년 12월 19일

ABSTRACT

In this paper, an ontology-based keyword structuring method is proposed to represent the knowledge structure of scholarly documents and to make inferences from the semantic relationships holding among them. The characteristics of thesaurus as a knowledge organization system(KOS) for subject heading is critically reviewed from the information retrieval point of view. The domain concepts are identified and classified by analysis of the information activities occurring in a general research process based on scholarly sensemaking model. The ontological structure of keyword set is defined in terms of the semantic relationship of the canonical concepts which constitute scholarly documents such as journal articles. As a result, each ontologically structured keyword set of a document represents the knowledge structure of the corresponding document as semantic index. By means of the axioms and inference rules defined for information needs, users can efficiently explore the scholarly communication network built on the semantic relationship among documents in an analytic way based on the scholarly sensemaking model in order to efficiently retrieve the relevant information for problem solving.

KEYWORDS

Knowledge Organization System, Thesaurus, Ontology, Semantic Relation, Inference, Semantic Indexing, Analytic Exploration

1. 서론

연구 · 학술문헌은 궁극적으로 특정 주제의 이슈 및 문제에 대한 연구자의 관점과 주장을 기술하는 정보이다. 정보이용자로서 연구자는 비판적 사고를 통해 타 연구자의 연구 성과를 이해 · 분석하고, 정보 생산자로서 파악된 문제점의 해결을 위해 새로운 아이디어를 뒷받침하는 다양한 연구성과를 참조 · 활용하여 도출된 새로운 연구 성과를 제시한

다. 연구 · 학술정보의 배포와 유통을 통해 이루어지는 학술커뮤니케이션은 다양한 연구주제에 대한 연구자간의 지속적인 상호 논쟁적 의사소통과정으로서, 연구자의 관점과 주장을 반영하는 각각의 연구 · 학술정보 간에는 자연적으로 다양한 형태의 '논쟁적' 의미 관계가 성립한다(Buckingham 2006). 연구자는 효과적이고 성공적인 '논쟁'을 위해 연구수행 단계에 따라 형성 · 조정되는 관점과 활용목적에 따라 다양한 성격의 연구 · 학술정

보를 필요로 한다. 연구자의 정보활동은 수요와 목적에 적합한 정보의 탐색과 해석을 통한 그 주제 영역에 대한 이해형성과정으로서, 정보검색은 수요에 적합한 정보의 접근과 획득을 뒷받침하는 핵심적 정보행위이다. 현행 정보검색서비스가 기초하고 있는 색인어 기반 정보검색 기술은 대상정보의 개념을 가장 적절하게 표현하는 어휘를 확률·통계적 기법에 따라 자동 추출 또는 해당 분야 전문가가 직접 추출하여 이용자에게 제공함으로써 수요에 적합하다고 정의되는 정보의 검색과 접근 경로를 제공한다.

최근 서지정보와 같은 메타데이터 모델과 시소러스 등 통제 색인어의 관계 구조화 등 해당 정보의 콘텐츠 속성 표현을 통한 정보 검색 단계 극복을 위해 용어의 개념 속성과 개념 간 관계 구조화를 지원하는 온톨로지의 도입·적용 방안에 대한 연구가 활발하게 진행되고 있다. 연구·학술정보의 접근성 향상을 위한 온톨로지 기반의 지식체계 구축방안에 대한 연구는 용어, 문서, 담화 등 3개 영역별로 구분할 수 있는데 현재 진행되고 있는 각 영역별 주요 연구주제를 살펴보면 ①시소러스 등 기존 전문 용어 의미 구조체계의 기능적 제한성 극복을 위한 온톨로지 기반의 관계유형의 세분화 방안 연구, ②연구·학술정보의 내재적 개념체계 및 논리구조의 체계화를 통한 정보 모델링 연구, ③연구·학술정보 간 성립하는 의미관계의 체계화를 통한 온톨로지 기반의 지식체계 모델링이 있다.

본 연구에서는 학술커뮤니케이션 관점에서 연구자의 의미기반 정보서비스 수요와 적합성 판단기준에 대한 이해를 통해 현행 시소러스 관련 연구의 한계와 문제점을 파악하고, 연구·학술정보의 개념구조 및 관계 분석을 통해 담화적 관계로 구성된 학술정보 의미 네트워크의 탐색과 분석을 지원하는 지식조직체계 구축방안을 제시하고자 한다.

2. 연구자의 정보활동 및 정보수요

연구프로세스 단계별로 수행되는 연구자의 주요 정보활동들을 학술커뮤니케이션 관점에서 살펴보면 다음과 같이 기술 될 수 있다.

첫째, 연구주제 및 아이디어의 발굴·수립 단계에서 연구자는 관심을 갖는 연구주제 분야의 기존 연구 성과에 대하여 비판적 사고를 통해 문제점 및 개선방안을 검토한다. 이 과정에서 새로운 접근방안을 제시하거나, 기존 결과를 개선·발전시킨 후속 연구, 또는 해당 연구 성과를 다른 분야의 문제에 적용한 응용 연구 등 다양한 성격의 관련 연구들에 대한 기초자료조사가 이루어지게 된다. 이를 통해 연구자는 해당 연구주제 관련 선행·관련 연구 정보의 수집과 이해를 통해 해당 주제 분야의 연구동향을 파악하고 자신의 연구의 목적과 가설과 같은 연구의 기본방향과 관점의 수립, 그리고 방법론, 적용분야 등 연구설계 초안을 수립한다.

둘째, 연구수행 단계에서 연구자는 연구계획 단계에서 수립한 목적, 가설, 연구방법론의 타당성과 적합성 그리고 완성도 확보를 위해 추가적으로 요구되는 정보를 탐색한다. 자신의 연구가설을 입증 증명해나가는 과정에서 그 가설을 뒷받침하거나 또는 자기 비판적인 시각에서 보완을 위해 관련 정보들의 종합적 이해와 선별을 통해 참조 및 활용함으로써 연구의 가설 및 방법론의 수정·보완 또는 새로운 아이디어 창출 등이 이루어지는 이해형성 과정이다.

셋째, 연구결과의 구조화된 문서 작성 및 배포 단계로서, 연구자는 자신의 연구 성과 즉, 특정 연구주제에 대한 연구자의 고유의 새로운 주장을 정형화된 문서 형식을 통해 작성하고 비공식 또는 공식적인 정보유통채널을 통해 배포·공유하게 된다. 이 과정에서 연구자는 앞서 수집·조사한 자료와 연구프로세스 단계별로 도출한 기초 연구데이터를 자신의 연구 성과에 대한 설득력 있는 주장과 명확한 이해를 위해 분류·선별하고 도입(introduction), 방법(methods), 결과(result), 결론(conclusion/discussion)으로 구성되는 일반적 연구학술 문서의 개념구조에 따라 각각의 정보들 간에 성립하는 유기적 의미관계를 논리적으로 표현하게 된다.

학술커뮤니케이션은 이러한 연구자의 연구 프로세스와 관련된 전주기적 정보활동을 통해 이루어지는 자연적인 의사소통 과정으로서, 결과적으로 연구·학술정보 간에는 논쟁적 의

미관계로 연결된 담화 네트워크(discourse network)가 형성된다(Buckingham 2000). 연구자는 특정한 주제영역에 대하여 새로운 아이디어와 연구성과물을 제시하고, 기존 연구와 관계정의를 통해 새로운 아이디어의 타당성, 차별성 및 중요성을 입증하고자, 학술 커뮤니케이션 네트워크상에서 해당 연구주제영역의 기존 연구·학술정보 간의 성립하는 관계를 파악하고 적합한 정보를 검색·활용하기 위해 다양한 목적의 정보활동을 수행한다.

문제 해결을 위한 이해형성 과정에서 연구자의 정보수요에 적합한 정보검색서비스 제공을 위해서 각각의 정보활동이 수행되는 상황(context)에 대한 이해가 필요하다. 연구자는 자신의 가설을 뒷받침하고 주장의 타당성과 완성도 확보를 위해 적합하고 신뢰할 수 있는 다양한 성격의 연구·학술정보 및 데이터를 필요로 한다. 예를 들면, ①연구가설의 근거와 필요성을 뒷받침 하는 이론, 사상, 법칙 등과 관련된 기초연구정보, ②연구 방법론의 타당성 및 적합성을 뒷받침하는 응용과 활용사례 연구, ③자신의 연구결과의 독창성 및 유효성 입증을 위한 비교분석을 위한 자료 등, 목적별로 다양한 관점과 주장을 담은 연구·학술정보들이 해당되며, 연구자는 그 정보들을 종합적으로 활용하여 자신의 가설과 주장의 논리체계를 구성한다. 연구수행 과정에서 단계별로 이루어지는 다양한 정보활동의 목적을 개념적으로 유형화하면 다음과 같다.

첫째, 해당 연구·학술정보를 통해 표현되

는 연구자의 관점과 주장이 속하는 해당 연구 주제 분야의 맥락 이해이다. 연구자는 어떤 특정 연구·학술정보의 획득 시 일반적으로 학술적 맥락(scholarly context), 즉 해당 정보와 의미관계를 갖는 다른 정보의 대상과 범위에 대한 이해를 통해 그 정보가 다루고 있는 연구 주제와 관련 동향 및 주제 도메인상의 개념적 위치를 파악할 수 있다.

둘째, 해당 연구주제 대한 관점의 다양성(scholarly perspective) 조사이다. 유사·동일한 연구 주제 및 연구문제의 해결을 위해 연구자는 고유의 아이디어와 연구자는 연구수행을 위한 기본 방향 설정 및 접근방법 검토를 위해 해당 연구주제를 다루고 있는 다양한 시각과 관점에 대한 종합적인 조사를 수행한다.

셋째, 해당 연구·학술정보의 학술적 영향성(scholarly impact) 측정 및 평가이다. 연구자는 참고문헌 등 직접적이고 명시적인 연관성 정보의 정량적 분석을 통해 측정되는 인용 지수를 통해 해당 정보의 영향성을 파악하고 대상 정보의 분석 및 이해의 우선순위 설정 등에 참조 할 수 있다. .

넷째, 해당 연구 성과의 학술적 일치성(scholarly consistency) 검토이다. 동일한 문제에 대하여 관점, 대상과 범위, 환경에 따라 각기 다르거나 상반된 결론이 도출될 수 있다. 그러한 정보의 존재 유무를 파악하고 그 중 의사결정에 유효한 관련 연구의 이해를 통한 연구결과 자기 검증은 완결성 확보에 필수적 이해형성 과정이다.

다섯째, 해당 연구·학술정보의 학술적 수렴성(scholarly convergence) 파악이다. 각기 다른 관점과 방법론을 사용하지만 유사·동일한 결과 및 결론을 제시하는 수렴현상 파악은 해당 연구주제에 대한 폭 넓은 시각과 심층적 이해를 위한 중요한 정보이다. 특히 최근 연구주제 복잡화에 따라 다학제 융·복합 연구가 활발하게 확대되고 있으므로 학술적 수렴성 파악을 위한 연구자의 정보활동의 비중이 높아지고 있다. 위와 같은 연구자의 정보탐색 활동 지원을 위해서는 아래와 같은 질의 유형의 성격과 특성에 대한 심층적인 분석과 이해가 필요하다.

- 이 연구결과를 토대로 수행된 후속·심화 연구 R은?
- 이 연구결과에 대하여 이의를 제기하는 연구 R과 그 논거는?
- 다른 관점과 방법을 통해 이 문제 P에 대한 해결방안을 제시한 연구 R은?
- 해당 연구주제에 대하여 이론 T에 근거하여 수행된 다른 연구 R1 또는 다른 연구결과를 도출한 연구 R2는?
- 연구방법 M을 다른 도메인 D, E의 문제 해결에 적용한 연구사례 R은?
- 자료 S의 연구를 위해 수행된 연구의 방법론 M에 대하여 수행된 기초연구 R은?
- 이 연구결과의 해석·검증 결과와 유사한 결과를 도출한 연구 R들은?
- 문제 P 해결을 위한 연구가 기초하고 있

는 이론 T1, T2에 대한 연구는?

현재 영향성(impact factor) 평가를 위한 인용 분석을 통해 정보 간의 상호 연관관계를 간접적으로 파악할 수 있는 지원하는 서비스가 제공되고 있으나 연구자들은 위와 같은 비교·분석적 성격의 정보탐색을 위한 도구나 지식조직체계의 지원을 거의 받지 못하고 있는 실정이다(Buckingham S., Motta, E., and Dominique J. 1999). 이러한 유형의 질문의 적합한 정보의 탐색과 제공을 위해서는 해당 연구·학술정보를 구성하고 있는 개념체계의 논리적 구조와 정보 간의 의미관계를 분명하게 파악할 수 있어야 한다. 그러나 현재 대부분의 학술연구정보 검색서비스가 채택하고 있는 색인어 기반의 검색서비스는 다양한 확률 통계적 기법에 대한 연구와 개발 노력에도 불구하고, 위의 열거한 질의유형에 대응하는 의미기반 서비스 제공에는 여전히 기술적 한계로 인한 제약을 받고 있어 분석적 정보수요를 충족시킬 수 있는 지식조직체계 개발 및 정보 서비스의 제공이 요구된다.

시소러스는 기본적으로 분류체계나 주제명 표목표로서 표현하기 어려운 개념간의 관계를 다차원적인 용어구조를 통해 표현함으로써 정보의 효율적 접근과 활용을 지원하는 가장 보편화된 지식조직체계이다(KOS; Knowledge Organization System). 용어사전은 주제어 색인 시 대표어휘를 선정하거나 의미에 기반한 용어 간의 관계나 구조를 보여줌으로써, 연구자가 정보검색 시 동형어의어와 의미적으로

유사한 주제 색인어를 선정할 수 있도록 해준다. 연구자는 키워드 검색을 통해 찾은 정보들의 적합성을 그 개념이 어떻게 다루어 졌는지 다른 개념들과의 의미관계를 통해 정의되는 용어관계구조를 파악함으로써 판단하게 된다. Aitchison(2000)은 시소러스의 역할을 문서정보자원에 대한 정보검색도구이자, 특정 주제분야를 이해하기 위한 일종의 의미지도(semantic map)로서 개념 간의 상호관계를 표시하고 용어 간의 정의를 이해할 수 있도록 지원하는 지식 조직체계로 정의하고 있다(Aitchison, J and Gilchrist, A., Bawden, D. 2000). 그러므로 시소러스의 디스크립터로 구성된 문서정보자원 각각의 주제 색인어 리스트는 대상 정보를 구성하는 주요 개념을 표현하는 용어들의 집합으로서 각각 개념을 단순히 독립적으로 열거하는 것이 아니라 그 용어간에 성립하는 유기적 의미관계를 통해 바로 그 해당 정보의 지식구조를 반영하는 것으로 확대 정의될 수 있다. 해당 용어들을 포함하는 주제 색인어 리스트가 그러한 문서의 지식체계를 구성하는 개념간 의미관계가 표현 될 수 있다면 또한 개별 지식구조의 특성에 따라 문서 계층에서 새로운 의미관계가 성립하거나 유도가 가능하게 되므로, 주제 색인어 리스트의 용어 간의 의미관계는 도메인 지식구조 이해에 핵심적 기반이 되는 지식조직체계라 할 수 있다.

본 연구의 목적은 시소러스 용어관계 구조의 개선과 관련된 다양한 연구들을 정보검색

을 지원하는 도메인 지식조직체계 구축 관점에서 분석하고, 특성과 문제점을 파악하여 연구학술정보의 지식구조를 반영하는 온톨로지 기반의 시소러스 구축방안을 제시하고자 한다.

3. 시소러스 관련 연구동향

문헌정보학에서는 연구자의 정보수요에 적합한 정보의 검색과 접근을 위해, 해당 정보의 개념 속성을 반영하는 핵심적 항목으로 구성된 메타데이터와 주제명 표목(subject heading) 기술 방안에 대하여 다양하고 지속적인 연구를 수행하여 왔다. 시소러스는 특정 주제영역에서 사용하는 용어와 이들 용어 간의 의미관계를 체계적으로 구조화한 통제 색인어휘 집합으로서, 해당 정보가 내포하고 있는 개념과 의미를 가장 적절하게 반영하는 주제명 표목의 선택을 위해 상·하위 및 연관관계 등 3가지 유형의 계층적 의미구조를 갖는다(한상길 1999). 그러나 시소러스는 용어의 의미를 개념 속성정의 및 그 값을 통해 표현하지 않고 다른 용어와의 상대적 관계를 통해 간접적 표현하므로 정확한 의미 파악이 어렵고, 용어의 다의성 및 관계구조의 단순성 때문에 개념간의 의미관계 파악과 구분에 한계가 있다. 따라서 수요에 맞는 정보의 효과적인 탐색을 위해서는 특정 제약조건의 설정과 다양한 의미 관계를 설정할 수 있도록 개념 속성 및 의미관계 유형의 확장이 요구된다. 정보수요 및 정보검

색 환경 변화에 따른 시소러스 구축방안 연구로는 첫째, 시소러스 용어 다의성에 대한 체계적 구분을 위한 패킷 기반의 용어 범주화 연구(Broughton, V. 2006; Tudhope, D., and Binding C.), 둘째, 시소러스 용어 관계구조의 일관성 및 표준화를 위한 용어 관계정의 규칙 및 제약조건에 관한 정의모델 연구(Sager, J. C. and Ndi-Kimbi A. 1995) 셋째, 메타데이터 기반의 시소러스 개념 속성 확장연구(Hunter, J. 2001) 넷째, 도메인 지식조직체계의 구축 기반으로서 온톨로지 기반의 의미관계 구조화 연구(Matthews, B. 2004) 등 해당 주제영역의 개념과 개념간 성립하는 다양하고 심층적인 의미관계의 식별을 위해, 언어학, 전산학 등의 여러 지식조직체계 설계 방법론의 도입·적용방안을 도출을 위한 연구가 활발히 진행되고 있다.

국내 연구로는 용어정의 기반의 시소러스 의미관계모델을 정보통신분야 전문용어에 적용한 연구(김태수 2001), 과학기술 분야 전문용어를 대상으로 패킷기반 용어 범주 및 용어 관계속성 유형 연구(황순희, 윤애선 2005), 온톨로지 기반의 시소러스 관계유형의 세분화를 법률용어 적용방안 연구(유영준 2005), 인문사회과학 분야 주제 색인어와 해당 정보의 메타데이터 결합을 통한 시소러스 확장방안 연구(이정연 외 2007) 등이 있으며 대부분 새롭게 제안된 시소러스 구축방법론의 실제 적용을 통한 타당성 검토 및 개선방안 제시에 초점을 맞추고 있다. 그 외에 지식조직체계로서 다

양한 정보수요의 지원 및 서비스 제공을 위한 다양한 개선 방안 연구들이 수행되고 있다.

본 연구에서는 각각 연구유형의 접근방법의 특성을 각각 도메인 지식구조체계 및 정보 접근성 관점에서 검토하고, 한계 극복 및 문제점 개선을 위해 연구학술정보의 지식구조를 반영하는 온톨로지 기반의 주제 색인어 및 시소러스 구축방안을 모색하여 보고자 한다.

3.1 패싯(facet) 기반 시소러스 용어 관계구조

용어를 기본 정보단위로 하는 시소러스의 기본적인 개념관계는 ISO 2788: 1986(E)에서 계층관계(hyponymy), 동등관계(synonymy) 및 관련관계(association)로 정의하고 있다. 시소러스를 기반으로 어떤 특정 주제 도메인의 개념 범주화와 의미관계 정형화를 통한 구조화된 지식 표현을 위해서는 먼저 시소러스의 구조와 특성에 대한 명확한 이해가 필요하다. 시소러스는 용어를 기본 단위로 하는 용어 중심의 관계 구조를 표현하는 지식체계이다. 시소러스는 “그 개념에 해당하는 적당한 용어를 찾는 데 사용하는 용어집”으로서 각각의 용어에 대하여 계층관계(BT/NT)와 동등관계(USE/USF), 그리고 연관관계(RT) 등 3가지 유형으로 다른 용어와의 관계를 포괄적으로 구조화하고 있다.

용어는 주제 분야와 같은 관점과 다른 용어와 관계에서 다양한 의미 속성을 가지며, 의미 속성에 따라 관계가 성립하는 대상 용어 및 그

용어와의 관계유형 또한 결정된다. 예를 들어 ‘배’는 각각 인체기관으로서 복부를 나타내는 배(abdomen), 교통·운송수단 중 해상 이동을 위한 배(ship), 그리고 식물의 열매로서 배(pear)’는 다양한 의미를 갖는데 해당 용어가 이중 어떤 개념범주 또는 주제 도메인에 속한 것이냐에 따라 용어의 의미관계 구조, 즉 BT/NT/(USE/UF)/RT관계를 갖는 대상 용어가 달라진다. 이러한 개념적 구분을 위해 분류 시소러스에서는 다양한 용어 범주화의 기준을 수용할 수 있도록 패싯 관계를 표현하는 관계어를 사용하고 있다. 패싯은 분명하게 정의되고 상호 배타적이며 또한 특정 범주 또는 주제와 관련된 집합적 관점, 속성 또는 특성을 의미한다(Mapple, A, 1999). 패싯 분류는 한 용어가 관점에 따라 다른 용어와 갖는 다양한 의미관계를 표현 할 수 있도록 다중 분류체계를 하나의 용어에 적용할 수 있는 의미 속성 분석 방식이다. 앞서 살펴본 개념 패싯에 이어 어의관계 패싯을 통해 정의되는 연관어 유형을 살펴보면 다음과 같다. 예를 들어 ‘온실’은 다양한 관점, 즉 ①재료에 따라 목조 온실, 알루미늄 온실 등, 또는 ②재배작물의 종류에 따라 화훼온실, 과수온실, 그리고 ③목적에 따라 실험용 온실, 가정 온실 등으로 분류될 수 있으며 일반적으로 분류 주제명과 용어 간에는 상·하위 개념관계가 성립한다고 정의된다. 패싯 개념의 도입을 통해 용어의 다의성을 범주화를 통해 체계적으로 수용하고 각 패싯 내에 특수한 관계를 반영하는 하위 범주화를 통

해 단순 관련관계를 의미적으로 세분화할 수 있다. 패킷 기반 시소러스에 대하여 제기되는 문제점은 다음과 같다(최석두 외 2006). 첫째, 해당 도메인의 지식을 정확하게 표현하기 위해 요구되는 적절한 패킷 세분화의 수준의 결정이다. 개념 체계의 범주와 하위 세분화는 사물과 현상을 보는 관점에 따라 결정된다. 시소러스는 용어 범주화 수준에 대한 일반적 기준을 제시하거나 조건을 한정하지 않는다. 따라서 같은 용어집합이라 하더라도 관점에 따라 각기 다른 논리 구조의 고유 개념체계가 존재하므로, 특정 관점에 따라 구축된 기존 용어체계의 개념 패킷 구조를 단순히 일반화거나 세분화함으로써 다른 여러 관점의 용어 의미 및 의미관계를 표현하는 것은 현실적으로는 물론 논리적으로도 한계가 있다. 둘째, 각기 다른 관점에서 구축된 시소러스는 관점의 유사도에 따라 결과적으로 포함, 중첩, 상보관계가 성립될 수도 있으나, 이는 논리적 관계가 아니므로 개념 간 매핑을 통한 용어 의미관계 공유나 단순 병합 시 동질적 의미구조를 보장하지 않는다. 셋째, 특정 패킷 관계 관점에서 개념 패킷은 상호 배타적 관계를 갖지만 경우에 따라 그 구분과 범위가 모호한 경우가 발생한다는 점이다. 예를 들어 방식-유형', 공간-위치', 물질-성분-재료', 상태-속성-현상' 등 관계 패킷어 자체에서 의미 중첩이 발생하는 경우로서, 용어에 패킷 할당 시 의미관계 충돌의 우려가 있다. 넷째, 패킷 관계명 간에 성립하는 의미관계의 근접성이다. 물질(기기, 재료,

부품)과 부분(부품, 성분, 요소)은 경우에 따라 서로 치환이 가능함에 따라 일관되게 적용되지 않을 경우 상·하위 관계가 바뀌는 경우가 발생하여 전체 해당 패킷을 통해 기술된 지식체계의 일관성 확보에 문제가 있다. 왜냐하면 패킷 지시어 자체가 의미하는 바가 명확해야 하는데 개념의 용어화(lexicalization)를 통해 정의된 패킷 지시어가 그 의미를 명확하게 구분하여 표현하지 못하거나 다른 패킷 지시어와 의미 중첩이 발생함에 따라 패킷 할당의 일관성이 보장되지 않기 때문이다. 또한 방식-유형'을 포괄하는 용어가 존재하지 않거나, 방식'과 유형'을 의미적으로 세분화 할 하위 패킷을 정의할 수 없으면 두 패킷에 의한 용어 범주화는 일관성을 보장할 수 없기 때문이다. 한편 의미중첩이 발생하는 개념 패킷을 구분하지 않고 상위개념으로 포괄적으로 범주화하는 것은 기존 시소러스 단순 관계구조로의 회귀를 의미하기 때문에 용어 간의 다의성과 모호성 해결을 위해서는 상보관계에 있는 개념들로 세분화된 체계여야만 한다. 그러나 이 경우도 특정 개념, 즉 피정의항을 표현하는 단일 용어가 존재하지 않거나 용어의 의미를 정의하는 정의항을 구성하는 용어의 조합을 통해 생성된 복합명사 역시 언어적 한계에 의해 해당 개념의 의미와 정확히 일치하지 않는 경우, 여전히 다른 패킷 지시어와 의미 중첩이 발생할 수 있다. 또한 개념을 표현하는 용어가 없을 경우 하위 패킷에 의한 세분화를 할 수 없으므로 다의성 또는 모호성이 발생할 수밖에

없다. 그러므로 용어기반의 개념체계는 본질적으로 자체 완결성을 가질 수 없는 불완전한 지식조직체계이다. 이러한 완결성 한계 때문에 시소러스 구축은 단순한 개념범주 및 관계 구조에도 불구하고 전문분야의 주제 색인어 전거통제사전 및 온톨로지 기반 지식조직체계 구축 등에 실제 활용할 수 있는 양적·질적 수준확보를 위해서는 여러 단계에 걸쳐 지속적인 조사·수집·분석·평가 등 많은 시간과 노력이 요구되는 작업이다. 목적에 맞는 시소러스 구축을 위해 일반 또는 전문용어사전으로부터 기본용어를 추출하고, 기존 시소러스를 활용하여 연계·통합·정제 과정을 거쳐 정보 생명주기에 따라 지속적인 갱신 및 유지 관리를 전문 인력의 투입을 통해 수행하는 것은 비효율적이며 수십만 건을 상회하는 용어에 대하여 일관성을 보장하는 것은 불가능하다. 따라서 일관성 있는 관계구조의 시소러스의 효율적 구축을 위해 다양한 자연언어처리 기법을 적용하여 특정 전문분야 대용량 문서 집합으로부터 자동으로 등록후보를 추출하고 주제적·형태적으로 관련이 있는 용어군을 만드는 군집화 등을 지원하는 구축도구 개발(김원중 외 2007) 및 생명주기에 따라 용어관계의 자동 형성과 일관성 및 유효성을 체계적으로 관리기능개발(정한민 외 2005)에 많은 노력과 연구가 꾸준히 진행되고 있다.

패킷 정의를 통해 개념범주의 정의 및 용어 관계를 세분화함으로써 정보 이용자가 색인어 선정 시 정보수요에 맞게 해당 용어 배의 다

의성 구분을 위한 조건을 예를 들어 식물' 개념범주로 한정하고 장소/지역' 관계 패킷으로 구분되어 의미적 연관관계를 갖는 나주 배'를 주제 색인어를 선정하여 적합한 정보를 효율적으로 찾을 수 있는 장점이 있다. 그러나 관계 중심의 시소러스의 용어의미관계 구조는 여전히 해당 개념과 관련이 있는 용어를 보여줌으로써 그 중 적절한 용어를 선택하는 데는 유용한 지식기반이 될 수 있으나 그 용어가 표현하는 해당 개념의 의미적 식별과 이해를 위해 요구되는 고유의 속성정의나 의미관계에 대한 제약조건을 직접적으로 표현할 수 없으므로 해당 개념의 의미를 구체적으로 기술하는데 근본적 한계가 있음을 알 수 있다. 예를 들어 조선시대 역사연구에 있어 중요한 의미를 갖는 조선 왕조의 가족 및 친·인척관계에 대한 지식을 시소러스의 용어관계구조로 기술한다고 가정하여 보자. 우선 의미관계 세분화 관점에서 보면 기존 시소러스 용어관계구조에서는 이종사촌, 증조할아버지 등 다양한 가족관계를 단순히 RT로 표현함으로써 그러한 의미관계의 구분이 되지 않는다. 패킷 기반의 시소러스 용어관계구조에서는 대상 지식구조를 구체적 표현을 위해 삼촌-조카관계', 이종사촌관계', 증조할아버지-증손자관계' 구체적으로 구분하여 명시하는 각각의 패킷 지시어 정의를 통해 단종'과 세조'간의 친조카-삼촌이라는 명시적 의미관계를 기술할 수 있다. 그러나 기존 배'의 사례와 살펴본 바와 같이 적합한 주제 색인어의 선정을 위한 의미구분의 목적을 중심으로

도입된 패킷 체계와 그에 기반한 용어의미관계 구조 분석을 통해 그러한 수준의 주제 도메인의 지식을 도출해 내는 것은 한계가 있다. 왜냐하면 조선 왕조의 가계도에 대한 이해를 위해 요구되는 가족 및 친인척 관계를 표현하는 개념체계와 단종' 또는 단종'이 표현하는 어떤 개념을 -예를 들어 조선시대 왕'- 포함하는 정보를 검색하는 데 있어 주제 색인어로서 적합한 용어를 선정하는 데 요구되는 용어의미관계는 그 표현 형식 및 개념구조가 각각 다르기 때문이다. 삼촌', 이모', 고모부' 등 가족 및 친인척관계를 나타내는 개념은 부모관계 및 형제자매관계 등 기본 관계와 성별에 의한 제약 조건의 조합을 통해 간단한 공리 정의와 기술을 통해 각각의 개념을 식별할 수 있는 명확한 의미 정의가 가능한 지식이다. 그러나 세조와 단종 두 용어의 관계를 친인척관계의 세분화를 통해 정의된 패킷 지시어를 통해 명시적으로 친조카-삼촌관계'로 기술하게 되면 모든 가족 관계 유형을 명시적으로 기술하기 위한 패킷 지시어를 정의해야한다. 그러나 연역적 추론 등 논리 연산을 통해 외조카 - 외삼촌관계'와의 의미적 구분이나, 친삼촌'과 숙모'의 상관관계 등과 같은 간단한 가족관계에 대한 이해 및 분석을 수행 할 수 없다. 또한 패킷 기반의 용어관계구조를 통해 표현되는 주제 색인어 간의 관계나 각각의 주제 색인어가 의미관계를 갖는 관련어 간의 상호 연관성을 표현할 수 없으므로, 해당 주제 도메인을 구성하는 개념 간의 유기적 관계나 지식체계의 표현과 이해에

제한이 있으므로 주제 색인어의 개념속성과 개념 간의 의미관계를 기술할 수 있는 표현 형식과 논리체계가 요구된다.

3.2 용어 정의기반 시소러스 개념 모델

기존 시소러스 관계구조는 용어 간의 관계구조나 범위주기를 통해서 간접적으로 용어의 의미를 기술하므로 해당 용어의 본질적 의미 파악이 어렵고 시소러스에 선정되는 용어의 범위와 대상을 정형화하기 어렵다. 따라서 특정 개념이 속한 개념체계와 그 체계 속에서 특정 개념을 고유하게 식별하고 용어를 분류하기 위해, 분석적 정의를 표준형식으로 하는 용어정의 모델 기반 시소러스 구축방안이 제시되었다(Sager, J. C. and L'Homme M. C. 1994)

시소러스 관점에서 정의는 어떤 개체가 지닌 다른 개체와 구분이 되는 개념적 특성(내포적 의미)을 밝혀준다. 이러한 논리적이고 분석적인 방법으로써 정의를 사용하는 주된 목적은 ①용어와 개념 간을 등식으로 규정하고 ② 독립된 정의를 통해 용어를 식별하고 ③개념의 의미를 제시하기 위한 것이다(Sager, J. C. 1990). 용어정의 모델은 용어의 의미를 데이터베이스 환경에서 효율적으로 이용될 수 있도록 본질적 개념 속성을 논리적 분석과 정의규칙에 따라 기술함으로써 정형화된 분석적 정의구조로서, 개념 속성과 속성 값 그리고 속성 값 영역의 제약조건으로 기술하고 있다

(Sager, J. C. and L'Homme M. C. 1994).
(〈표 1〉 참조).

용어정의를 시소러스 관계구조에 도입하는 주된 이유는 특정한 추상화 수준에서 개념을 기술하고 그 개념을 관련된 개념과 구분하며, 개념체계에서 특정개념의 위치를 결정하기 위해 그 개념과 다른 개념과의 관계를 설정하고 개념의 범위를 한정함으로써 용어의 이용을 표준화하기 위한 것이다. 〈표 1〉에서 보는 바와 같이 용어정의 모델은 각 용어의 서술형 정의를 ①주제 분야, ②피정의항의 개념범주, ③정의하는 개념이나 정의항, ④정의항의 개념범주, ⑤피정의항과 정의항의 관계, ⑥피정의항과 관련 개념간의 본질적인 구별특성(중차, differentia), 그리고 ⑦기타 특성 등 총 7개 항목에 따라 본질적 속성으로 분해·구조

화하고 각 속성의 값에 대한 정의규칙을 제시한다. 지식조직체계 관점에서 볼 때, (1)번 항목은 주제 도메인을 명시하며 (2)번 항목과 (3),(4)번 항목은 해당 용어가 속한 개념 범주와 논리적 관계를 갖는 용어 및 그 용어의 개념범주의 정의 (5)번 항목은 용어의 정의방식에 따라 성립하는 용어간의 관계유형 정의 (6)번 항목은 피정의항 내에 구분이 되는 의미 속성에 따라 관련된 용어와 개념범주 그리고 관계유형을 정의하고 있다. 이는 각각 ⑤번 ⑥번 항목을 통해 정의된 용어 간의 관계구조를 각각 시소러스 관계구조의 계층관계 및 연관관계로 매핑함으로써 용어로 표현되는 개념의 본질적 속성을 시소러스 관계구조에 반영하고 세분화된 관계 패킷을 통해 관련어와의 관계 유형 구체화가 가능하다.

〈표 1〉 Sager와 L'Homme의 정의모델

피정의항(용어) = (1) 주제분야 (2) 피정의항의 개념범주 (3) 정의항(최근접유개념어) (4) 정의항이 개념범주 (5) 피정의항과 정의항의관계 (6) 피정의항과 관련개념을 구별하는 특성, 즉 중차 ① 본질 ② 목적 ③ 수단 ④ 상태변화 ⑤ 유사 ⑥ 위치 ⑦ 반대 (7) 이용자에게 유용할 것으로 보이는 비본질적 특성

용어 시소러스의 의미관계는 명제적 지식의 표현 형식으로서 용어정의 모델에서 개념 간의 관계는 계층관계, 분할관계, 대립관계, 기능관계 등 네 가지로 그 실제적 관계로 유형을 정의할 수 있다(Sager, J. C. 1990). 계층관계를 이용한 정의는 개념을 나타내는 상위어와 하위어의 논리적 배열로, 분할관계를 이용한 정의는 개념을 대상의 구성요소로서, 대립관계를 이용한 정의는 개념이 부정 또는 반대로 정의되는 것이고, 기능관계를 이용한 정의는 전체개념의 요소 및 특성을 포함하는 경우를 의미한다. 또한 연관관계는 등가관계에 포함되지 않으나 개념적으로 밀접한 관련이 있다고 판단되는 용어로 정의되어 있어 명확하

고 구체적 기준설정이 어려운데, 용어정의 모델은 관계어를 용어의 정의나 설명에서 필요한 용어로 한정함으로써 대상과 범위를 구체화 할 수 있다. 용어정의 모델을 도입한 시소러스 개발 및 구축을 위해서는 단계별로 다음

과 같은 절차적 특성과 개발 방법에 대한 고려가 필요하다. 첫째, 용어정의 모델은 기존 시소러스의 용어 관계구조를 대체하는 새로운 형식이다. 개념의 의미구조를 반영하기 위해서는 시소러스의 용어 관계정의가 한계가 있

〈표 2〉 용어의 내포적 정의 분석

피정의항	milk		cheese		yogurt	
피정의항의 범주	구체적 실체		구체적 실체		구체적 실체	
정의	Milk is an opaque white liquid produced by the mammary glands of female mammals (including monotremes)		Cheese is a food made from milk, usually the milk of cows, buffalo, goats, or sheep, by coagulation. The milk is acidified, typically with a bacterial culture, then the addition of the enzyme rennet or a substitute (e.g. acetic acid or vinegar) causes coagulation, to give "curds and whey"		yogurt is a dairy product produced by bacterial fermentation of milk. Fermentation of the milk sugar (lactose) produces lactic acid, which acts on milk protein to give yoghurt its texture and its characteristic tang. Soy yogurt, a dairy yoghurt alternative, is made from soy milk.	
정의항	liquid		food		dairy product	
피정의항의 범주	구체적 실체		구체적 실체		구체적 실체	
정의항과 피정의항의 의미관계	도메인		도메인		도메인	
특성	본질	opaque white	본질	curd	본질	lactose, lactic acid, milk protein
	근원	female mammals	근원	milk	근원	milk
	수단	mammary glands	수단	acidification, coagulation, bacterial culture, enzyme rennet	수단	bacterial fermentation
	유사					soy yogurt

으므로 용어의 의미정의 모델로 개념 의미구조를 표현하고 이를 다시 시소러스 구조에 반영시킴으로써 일정수준의 일관성을 확보하고 있다. 정보통신분야 전문 용어 시소러스(김태수 2001) 교육학 용어 시소러스(Sager, J. C. 1990) 등의 사례에서 살펴본 바와 같이, 용어정의 모델의 개념구조와 시소러스 용어관계구조(BT/NT /RT)를 구분하여, 전문용어사전의 지식구조를 반영하는 용어정의모델을 시소러스 용어관계구조의 일관성을 확보하기 위한 도구 및 자원으로서 활용방안을 제시하였다. 한편 용어정의 모델은 개념속성 정의를 통해 용어의 본질적 의미를 표현하므로 각 용어의 미관계 구조간의 의미관계 정의를 통해, 논리적 연관관계에 대하여 일정 수준의 추론이 가

능할 수 있다(고영만 2006).

〈표 3〉은 cheese와 yogurt가 각각 같은 근원(source)인 Milk를 응고 및 발효 등 각각 다른 방법의 가공을 통해 만들어진 식품으로서 의미적으로 볼 때, 근원을 기준으로 동일한 개념범주에 포함될 수 있음을 파악할 수 있으며 마찬가지로 다른 속성 및 속성들의 조합에 따라 개념범주를 정의할 수 있다. 다음은 정형화된 개념의미구조의 속성을 기준으로 개념범주를 정의할 수 있음을 보여주는 사례이다.

- 1) If the source of X is Milk and produced by coagulation with acidification or bacterial fermentation, then X is dairy product.
- 2) If X is produced by acidification

〈표 3〉 용어의 개념적 정의 분석

피정의항	인식론		경험론		합리론	
피정의항의 범주	철학		철학		철학	
정의	인간 지식의 근원, 본질, 방법, 한계를 탐구하는 철학 연구		모든 지식은 지각 경험을 통해서만 인식된다는 철학 이론		모든 지식은 경험이 아닌 이성을 통해서만 습득된다는 철학 이론	
정의항	연구		이론		이론	
피정의항의 범주	추상		추상		추상	
의미관계	도메인		도메인		도메인	
특성	본질	탐구	본질	주장	본질	주장
	목적	이해	목적	-	목적	-
	대상	지식	대상	지식	대상	지식
	속성	근원, 본질, 방법, 한계	속성	-	속성	-
	수단	-	방법	경험	방법	이성
	유형	경험론	도구	지각	도구	본유관념
		인식론				
반대	-	반대	-	반대	경험	

with bacterial culture or bacterial fermentation, then X is fermentation food.

위의 공리적 사실을 통해 cheese와 yogurt 를 포함하는 각각의 정보가 모두 diary product 범주에 속하는 개념을 다루고 있음을 유도할 수 있다. 패킷 기반의 용어구조는 의미 관계를 일정수준 세분화 할 수 있으나 관계구조의 일관성이 없어 위와 같은 완결성을 보장하는 개념 범주를 정의할 수 없다.

둘째, 개념들은 종차(differentia specifica), 즉 내포적 정의 외에 다른 다양한 맥락에서 정의적 관계를 가질 수 있다. 용어정의의 유형은 목적에 따라 기능적으로 나뉜다. 개념간의 관계유형을 정의모형별로 살펴보면 각 기능과 목적에 따라 각각 개념적 정의(conceptual definition), 조작적 정의(operational defini-

tion)와 이론적 정의(theoretical definition), 외연적 정의(extensional definition)등 다양한 형태로 구분된다. 개념적 정의는 그 개념을 표현하는 다른 개념의 논리적 관계(예: 무게: 특정 대상에 작용하는 중력의 힘), 조작적 정의는 관찰 불가능한 추상적 개념을 측정 가능한 구체적 현상으로 연결(예: 암페어: 초당 흐르는 전류의 양을 측정하는 단위), 이론적 정의는 구체적 관찰이나 경험 등 귀납적 추론에 의해 도출된 정의(예: 인플레이션: '초과수요', '비용인상', '관리가격통제' 등 물가를 결정짓는 요소의 불안정한 상태로 발생하는 지속적인 물가상승 현상), 외연적 정의는 해당 개념에 속하는 모든 개체의 열거(예: 2002년 축구월드컵 출전선수단: 박지성, 안정환, 유상철,...히딩크)등이 있으며 이들 용어정의 형식은 내포적 정의와 다르므로 해당 용어정의사례의 분석을 통해 피정의항과 관련 개념 간의 구분속성 등

〈표 4〉 시소러스 용어 클래스와 속성

까뮈(Camus, Albert)	
인명 - 정의	프랑스의 소설가(1913-1960), 평론(시지프의 신화)에서 부조리의 철학을 논하여 실존주의를 더욱 심화시켰고, 전후의 사상과 문학에 크게 영향을 끼쳤다. 1957년에 노벨문학상을 받았다. 작품에 <이방인>, <페스트>등이 있다. <표준국어대사전>
인명-생물년	1913-1960
인명-이칭/이형	까뮈[Camus, Albert], 까뮈,알베르[Camus,Albert] 알베르 까뮈[Camus,Albert] 알베르 카뮈[Camus,Albert] 카뮈,알베르[Camus,Albert]
인명-출생지	알제리 몽드비
인명-지명	소설가, 극작가
인명-지위	
인명-소속기관	
인명-저작	결혼(산문), 계엄령(희곡), 독일인에게보내는편지(편지), 시지프의 신화,오해(희곡), 이방인(소설), 전략(소설), 페스트(소설) 표리(소설)
인명-국적	
인명-시대	

각각의 유형에 적합한 정의모델 개발이 요구된다. 또한 개념에 따라 다양한 범위의 설명이 요구됨으로, 용어사전의 정의에 대한 구분적 분석에 따른 용어간의 관계유형 정의에 의한 의미표현의 일반화에는 일정한 한계가 있다.

〈표 4〉는 다양한 용어정의를 참조하여 용어 정의모델에 근접한 사전적 정의를 대상으로 추상적 개념의 의미표현에 적용한 사례이다. 개념적 용어정의 모델을 통해 표현되는 '경험론'과 '합리론'의 의미구조가 두 개념이 동일한 도메인에서 본질과 대상이 동일하나 방법과 도구 속성에서 대립적 관계를 갖는다는 것을 나타내고 있다. 각 주제어를 포함하는 연구학술정보는 해당 도메인에서 이론적 관점에서 대립적 관계가 성립하는 것을 부분적으로 유도할 수 있다.

셋째, 용어정의 모델에 기반한 시소러스 구축의 경우 고유명사에 대한 처리방안이 필요하다. 인문사회과학 분야에서는 춘향전, 모택동, NATO, 석굴암, 레오나르도 다빈치, 경주, 9.11 테러사건 등 다양한 고유명사가 주제 색인어로 사용될 수 있다. 만약 이들 고유명사를 시소러스 디스크립터에 포함시킬 경우, 식별기능(designator)을 갖는 고유명사의 특성 상 전문용어 중심의 내포적 정의방식과는 다른 별도의 디스크립터 정의모델이 필요하다. 고유명사의 경우 해당 고유명사의 개념 범주(인물)에 대한 고유 정의모델을 정의할 수 없으므로, 각 고유명사와 개념-사례 관계를 갖는 해당 개념의 용어정의 모델에 따라 정의항과 피정의항의 개념범주 및 관계구조를 갖

는 상속받는 체계가 의미 일관성 유지 차원에서 타당한 것으로 판단된다. 예를 들어 '춘향전'과 '문학작품'의 관계가 개념-사례에 해당된다고 할 때, '춘향전' 자체의 용어정의 모델을 갖는 것이 아니라, '문학작품'의 내포적 정의에 의해 요구되는 7개의 항목을 충족시키는 정의모델로서 의미가 기술될 수 있도록 정형화하는 것이 필요하다. 또한 용어정의의 경우 같은 용어라 하더라도 전문 분야별로 다른 용어관계구조를 갖게 되므로 해당 용어정의가 유효한 주제영역을 해당 용어의 속성 값으로 정의하고 이에 대한 구분이 필요하다.

기존 시소러스와 대비되는 용어정의 모델 기반 시소러스의 특징은 해당 주제 분야의 개념이 전문 용어사전과 같은 바로 그 분야에서 인정되고 있는 지식 구조를 직접 반영하고 있다는 점과 그 개념이 속한 지식분야의 다른 모든 개념과 관련하여 그 개념의 위치를 한정할 수 있고, 이에 따라 디스크립터를 선정하는데 일관성을 유지 할 수 있다는 점이다. 또한 제시된 7개의 내포적 특성을 용어관계에 구체적으로 반영할 경우, 표현과 구조가 일관성을 갖는다는 것을 가정할 때, 그를 기준으로 특정한 공리를 정의함으로써 다양한 개념 범주를 새롭게 정의할 수도 있다. 그러나 이러한 사실은 역설적으로 기존 시소러스 용어 관계구조 자체로는 개념 의미구조 및 관계를 일관되게 표현할 수 없으며, 종차 등 어떤 개념 모델을 기반으로 의미를 일정한 형식에 따라 기술하는 정보, 즉 전문 용어사전, 전문정보 데이터베이스

스, 어휘망, 주제분야 온톨로지 등 다양한 형식의 구조화된 정보가 필요하다는 것을 의미한다. 전문 용어사전의 정의를 분석하여 정의 모델 형태로 가공하고 또한 시소러스의 용어 구조를 수정하거나 확장하는 것 역시 많은 전문 지식과 노력이 요구되기 때문에 최근 들어서는 세분화된 의미관계의 일관된 표현을 위해 기존의 구축된 다른 구조화된 정보자원의 활용 및 정보모델링 기술을 적용하여 개념의 범주화 및 속성 정의 문제를 해결하고자 하는 방안이 추진되고 있다(최석두 외 2006).

3.3 메타데이터 기반 시소러스 용어관계 확장 모델

메타데이터는 대상 객체의 속성이나 내용을 특정 관점에서 기술하는 데이터로서, 예를 들

어 문헌의 서지데이터, 인물의 프로파일, 제품의 명세 등이 여기에 해당되는데 다양한 메타데이터가 분야와 목적에 따라서 시소러스의 관계 확장 및 범주화를 위한 기반 정보로 활용될 수 있다. 앞서 살펴 본 용어정의 모델 도입을 통한 시소러스 개발연구와 차이점은, 전자의 경우 해당 지식구조를 시소러스 용어구조로 변환한 반면, 기존 시소러스 용어관계구조와 용어정의모델, 그리고 메타데이터 데이터 요소를 직접 도입한 개념구조를 확장' 시소러스로 정의하고 있다. <표 5>는 표준국어대사전에 나온 용어의 서술 내용을 토대로 각 속성을 정의하고 관련 데이터베이스의 데이터요소 항목을 수집 포함하여 정의한 클래스 및 속성으로 표현된 개념구조이다(이정연 외 2007). 항목명인 인명-정의'에서 인명은 클래스를, 정의는 속성을 나타낸다.

<표 5> 1차 및 2차 개념관계

까뮈(Camus, Albert)		
UF	Albert Camus	프랑스어명/이름
	Camus Albert	프랑스어명/이름
	까뮈[Camus, Albert]	음차명/이름
	알베르 까뮈[Camus, Albert]	음차명/이름
	알베르 카뮈[Camus, Albert]	음차명/이름
	카뮈 알베르[Camus, Albert]	음차명/이름
RT	노벨 문학상	상/수상자
TITLE	결혼(산문)	산문/저자
	계엄령(희곡)	희곡/저자
	독일인에게 보내는 편지(편지)	편지/저자
	반항적 인간(평론)	평론/저자
	시지프의 신화(평론)	평론/저자
	오해(희곡)	희곡/저자
	전략(소설)	소설/저자
	페스트(소설)	소설/저자
	표리(소설)	소설/저자

최석두(2006)는 바람직한 것은 메타데이터의 어느 요소에 포함되어 있는 것이라도 그것이 용어로서 의미가 있는 것은 모두 시소러스에 포함시키는 것이다”라고 확장’ 시소러스의 범위를 서술하고 있으며 이정연(2007)은 “그러나 정의, 출생지, 직업 등에 포함되어 있는 용어나 개념과 모두 반드시 관계를 가져야 한다는 것을 의미하는 것은 아니다”라고 기술하고 있다. 그러나 이와 같은 용어 중심의 개념속성 구조화 방식은 지식조직체계 구축 및 전거통제어휘집 두 관점에서 그 타당성을 고려하여야 한다.

〈표 6〉에서 왼쪽열은 소위 말하는 1차 개념 관계를 나타내는 관계명으로 오른쪽 열은 2차 개념관계로서 각각 1차 개념은 개념의 속성을, 2차 관계는 개념간의 관계를 각각 나타낸다.

용어와 관련된 메타데이터 항목으로 시소러스의 개념 속성 및 용어관계를 확장하는 경우 역시 고유명사 및 다른 주제 색인어의 구분이 필요하다. 고유명사의 경우, 용어정의 모델에서 검토된 바와 같이 용어 의미구조 차원에서 정규화 된 개념모델 정의가 일반 주제 색인어와 구분이 되므로 위의 접근방법을 단순하게 그대로 적용하거나 일반화하기는 힘들다. 반면 일반 주제색인어의 경우, 용어와 관련된 구조화된 정보자원으로부터 의미관계가 성립하는 용어를 일정한 기준에 의해 선정하여 용어관계 구조를 정의하므로 해당 정보자원의 지식구조를 반영하게 된다. 대상이 특정 주제 분야 전문 용어사전일 경우, 해당 시소러스는 다른 주제 분야와 구분이 되는 다른 의미관계 구조를 갖게 되므로 전거 통제 시 그 주제영역을 명시

〈표 6〉 AGROVOC relationships compared with more differentiated relationships of a Food and Agriculture Ontology

AGROVOC	Food and Agriculture Ontology
Undifferentiated hierarchical relationships	Differentiated relationships
milk	milk
NT milk fat	includesSpecific> cow milk containsSubstance> milk fat
cows	cows
T cow milk	hasComponent> cow milk*
Cheddar cheese	Cheddar cheese
T cow milk	madeFrom> cow milk
	Rule 1
	Part X <mayContainSubstance> Substance Y
	F Animal W <hasComponent> Part X
	ND ngests> Substance Y
	Rule 2
	Food Z <containsSubstance> Substance Y
	F Food Z <madeFrom> Part X
	ND Part X <containsSubstance> Substance Y

함으로써 정확한 용어 선정이 이루어 질 수 있어야 한다. 그러나 위의 정의된 바와 같이 용어와 관계된 메타데이터에서 일부 항목을 시소러스 용어구조에 선별적으로 반영을 하게 되면 그러한 구분을 할 수 없으므로 주제 색인어 전거통제 및 검색 시 특정 주제 영역에서 유효한 개념의 의미구조를 선택하여 검색을 할 수 없다. 이를 위해서는 해당 용어가 표현하는 개념 의미구조가 유효한 영역을 정확하게 명시할 수 있는 분류체계와의 연계를 통한 식별이 필요하다. 또한 의미적 상호 운용성을 위해 'TITLE'과 같은 관계속성에 대한 의미 정의와 해당 속성에 적합한 값의 영역에 대한 정의 역시 필요하므로 두 개의 상이한 구조와 성격의 지식조직체계의 연계를 위해 기존 정보 모델의 확장을 통한 물리적 통합으로는 문제 해결에 한계가 있으므로 새로운 지식조직체계의 구축을 위해 요구되는 조건을 고려하여 정보 모델의 설계가 필요하다. 따라서 메타데이터기반의 시소러스의 확장의 경우, 현재 다양한 지식조직체계의 상호 운용성 및 연계 통합과 관련하여 진행되고 있는 ISO 11179(Metadata), ISO 704(Terminology), ISO 12620 (Data Categories) 등 관련 표준 전문가들 간의 논의와 연구(Bargmeyer, B. 2005) 등의 결과를 고려하여 체계적인 정보 모델 수립에 향후 면밀한 연구가 요망되며 대상 정보의 지식체계를 반영하는 구조화된 주제 색인어 리스트의 정보 모델 및 구축을 위해서는 문서구조를 표현하는 상위 온톨로지와 연동이 필요하다.

3.4 시소러스와 온톨로지의 연계 활용방안

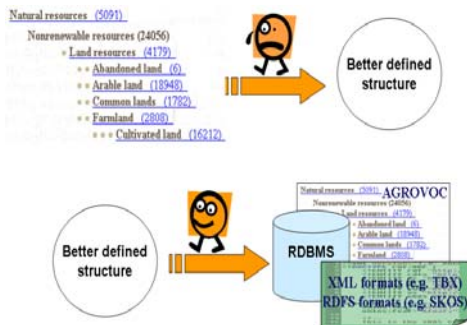
정보 모델 관점에서 시소러스의 용어관계구조와 비교하여 볼 때, 도메인 온톨로지는 특정 주제 관점에서 용어를 포함 각 계층을 표현하는 정보들이 표준화된 형식체계에 의해 유기적으로 결합된 지식조직체계이다. 도메인 온톨로지는 도메인에 내재한 개념요소의 명시적 정의나 개념 간의 의미관계 및 공리체계를 정의하여 지능적인 정보처리를 지원하는 지식기반을 제공하므로, 의학(NCI) 분야(Ceusters, W. and Smith B. 2005) 등 여러 전문 분야의 온톨로지 구축이 활발하게 추진되고 있다. 도메인 온톨로지 구축은 복잡한 정보처리를 위해 요구되는 정보의 질적 수준의 확보를 위해 많은 인력과 자원이 소요되는 중·장기적 과제로서, 기술적 이슈와는 별도로 구축 효율성이 매우 중요한 관건이 되고 있다. 그러므로 시소러스와 같은 기 구축된 지식조직체계의 확장 또는 상호 운용성 확보를 통한 지식조직체계의 통합적 연계 활용방안에 대한 연구가 이루어지고 있다. Soergel(2004)은 UN 식량농업국(FAO)의 다국어 시소러스인 AGROVOC의 분석을 통해 온톨로지 개념 간의 관계설정 에 대한 연구를 수행하였는데, 특정 도메인의 용어 시소러스에 함축되어 있는 개념, 용어, 문자열 계층의 관계를 구분하고 시소러스에 표현된 개념관계를 온톨로지의 기술형식 및 개념 구조화 방법을 <표 7>과 같이 기술하고 있다(Soergel, D. 2004).

그러나 실제로는 상대적으로 단순한 구조의 정보자원의 확장을 통해 다양한 정보처리를 지원하는 개념속성과 의미관계로 구조화된 지식조직체계의 변환 및 정의 작업은 시소러스 고유 구조적 형태적 특성의 유지는 물론 새롭게 구성되는 지식조직체계의 구조적 일관성 확보가 매우 어렵다(〈그림 1〉 참조). 더구나 기존에 구축된 시소러스의 용어체계가 특정 정보활용지원을 위해 요구되는 도메인 온톨로지의 개념체계를 전적으로 반영하는 경우는 거의 기대하기 어려우므로 시소러스를 구성하는 개별 어휘단위의 수정 작업을 통해 도메인 온톨로지 구축을 하는 것은 자원 구축의 효율성 및 질적 수준을 동시에 고려해야하는 점을 감안할 때 현실적으로 매우 어려운 과제이다. 그러한 점들을 고려하여 FAO는 식량 농업관련 개념 기술을 위해 새로운 온톨로지 개념모델을 설계하고 기존 관계형 DB형식의 시소러스의

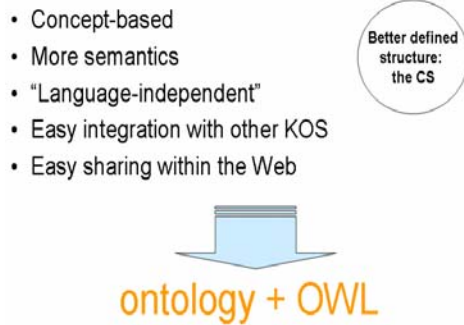
개념분석을 통해 변환 가능한 부분을 새로운 OWL-DL형식으로 재구성(re-engineering)하였다(〈그림 2〉 참조).

온톨로지의 질적 수준의 확보를 위해 전체 구조를 재검토하고, 계층구조의 정비, 관계속성의 세분화, 상세한 어휘정보의 추가, 의미정의 등을 기술함으로써 다양한 목적의 정보처리 수준에 맞는 지식조직체계로 변환을 지원하는 개념 서버를 구축·제공하고 있다(Fischer, D. 1998)²⁾.

Fischer(1998)는 주제 색인어 전거통제 및 정보검색을 위한 어휘집으로서 구축된 시소러스의 용어관계구조를 사실 서술(fact description) 및 추론을 위해 구조화된 온톨로지 로 변환 및 확장하기 위해서는 BT로 표현된 모든 계층적 포함관계(subsumption)에 대한 세분화 및 재정의와 더불어 개념과 메타개념의 관계와 구분이 필요함을 ISO 2788, WordNet,



〈그림 1〉 FAO 지식조직체계



〈그림 2〉 OWL기반 용어 온톨로지

2) 현재 FAO 온톨로지 기반의 Agricultural Ontology Service Concept Server(AOS/CS)를 구축함으로써 시소러스, 메타데이터 스키마, SKOS, TBX와 후방호환성을 지원한다(<http://www.fao.org/aims/aos.jsp>).

Cyc의 지식구조와 상세한 상호 비교 분석을 통해 입증한 바 있다. 또한 주제 색인어 전거 통제 및 정보검색을 위한 실용적 관점의 시소러스의 용어관계구조와 개념속성 정의 및 추론을 위한 논리적 관점의 도메인 온톨로지의 목적과 기능에 따른 차이와 특성을 비교하고 온톨로지가 전거통제 및 의미기반 정보처리를 지원하는 폭넓은 응용범위를 갖는 지식기반으로서 역할을 전망하였다. 그러나 전거통제 및 키워드 기반 정보검색이나 의미기반 분석 서비스에 도메인 온톨로지를 본격적으로 구축·활용하고 있는 사례는 아직 찾아보기 힘든 상황이다.

한편 시소러스의 용어관계구조의 자체 확장 방안 연구 외에 동일 도메인에 존재하는 이질적 구조의 분산 정보자원의 상호 운용성 확보 및 통합검색 지원을 위해 각 메타데이터 용어의 의미관계를 표현하는 시소러스 구축방안(Hunter, J. 2001) 및 기존의 시소러스가 기능적 독립성을 유지하면서 도메인 온톨로지와의 연계하여 통합검색을 지원하는 RDF기반의 개념체계(SKOS) 연구(Matthews, B. 2004) 등이 있다. 메타데이터 용어 간에 성립하는 동형어, 반의어, 상·하위 및 연관관계를 파악하여 분산 정보자원에 대한 통합적 의미 검색 또는 용어관계구조를 참조하여 해당 개념을 포함하는 정보 간의 관계 추론 기능을 제공할 수 있다. 국내의 경우 이러한 선행 연구들을 바탕으로 검색 효율성 관점에서 시소러스의 용어관계를 표준 온톨로지 형식언어(OWL)의

미관계를 통해 세분화함으로써 구문적 상호 호환성 확보와 다국어 정보에 대한 통합 검색 효율성 제고(조현양, 남영준 2004), 한국 역사 용어시소러스를 대상으로 시소러스의 온톨로지 변환을 지원하기 위해 SKOS를 기반으로 단계적 변환 알고리즘 및 구축방법 등에 대한 연구(한성국, 이현실 2006)가 수행된 바 있다.

이상 지금까지 검토한 의미 기반의 정보탐색 및 접근 지원을 위한 시소러스 용어관계구조화 및 의미관계 세분화 관련 연구의 공통적 문제는 용어의 의미관계 구조화가 근본적으로 검색에 적합한 주제 색인어의 선정을 위한 용어관계구조 기술에 한정되어 있다는 점이다. 정보 이용자는 어떤 개념을 표현하는 특정 용어에 대하여 세분화된 용어관계구조를 통해 관련된 용어를 파악하고 그 중 자신의 정보수요에 가장 적합한 용어를 선택함으로써 효과적인 검색 결과를 얻을 수 있지만 이러한 용어관계구조를 통해 해당 용어가 표현하는 개념이나 그 개념을 다루고 대상 주제 도메인의 지식구조에 대한 이해는 매우 제한적일 수밖에 없다. 주제 색인어 전거통제 및 검색지원 도구로서 시소러스의 고유 기능을 유지하면서 그러한 문제점을 보완하기 위해 용어정의모델 도입 및 해당 용어관련 DB의 메타데이터와 연동 방안이 제시되었다. 전문용어사전 또는 관련 정보 DB 등 정형화된 지식자원과 변환 및 융합 등 연동을 통해 정형화' 및 확장'된 용어 의미관계구조는 검색하고자하는 개념에 적합한 주제 색인어의 선정을 통해 구체적인 특정

〈표 7〉 인문사회과학 분야 학술논문의 구성 개념 범주

개념 범주	개념범주 패시이 할당된 용어의 예
1. 개념/정의	
개념/정의	정보문해/남녀평등/유비쿼터스/매트로섹슈얼/온톨로지
2. 이해	
이론/원리/법칙/모델	경험론/계임이론/평등원리/파레토의 법칙/롱테일 법칙/카노모델/피그말리온효과/FRBR
3. 관점	
학제/학파/주의	고전문학/성리학/언어심리학/공리주의/법다원주의/해체주의
4. 구성요소	
인물	박정희/헤겔/이순신/우사인 볼트
그룹/단체/조직/기관	수니파/경실련/삼성생명/서울중앙우체국/국회
공간/장소	국립중앙박물관/국회도서관/서울숲공원/한국
시간/시점/시대/시간	2008년/현재/조선시대/산업화후기
제품/기술/시설/물질/시스템	전자화폐/PMP/IPTV기술/국가전산망/황토/정보네트워크
자연/환경	동물/식물/곤충/산/바다
사건/행동	911테러/6.4한미정상회담/5.18민주화운동/쓰촨성 대지진
콘텐츠	춘향전/밀레의중(그림)/배트맨/월광 소나타/불국사/축구/뮤지컬
5. 체제	
법/제도/교육/의식/종교/문화	친일재산환수법/국가보안법/종합부동산세/주5일제근무/전송규칙/윤리의식/이슬람교/음주문화/결혼문화/중국문화
6. 문제 · 이슈	
정치 · 외교	장기집권/국회파행/독도영유권/한미관계악화/동북공정
경제	경기침체/물가불안/고유가/실업률 증가
사회	인권침해/입사파열/수월성교육/사회양극화
역사	친일반민족행위/을사조약
문화 등	사회주의 리얼리즘/문학의 사회참여
7. 특성	
감정/심리	불안감/기대심리/색채심리민족의식/잠재의식/공동체의식
증상/증후	조급증/우울증/새집증후군
성질/성향	개방성/동성애/보수/진보/중도/무자비/폭력성
상황/상태	익명성/합법성/정당성/유사성/직업
속성/능력	견고/공간지각력/충격흡수력/독서능력/학습력
8. 방법	
조사분석방법	역학조사/심리테스트/설문조사/인터뷰/통계조사/자동변환
9. 수단/도구	
수단/도구	다변량분석/T-검증/표현형질문/알고리즘
10. 데이터/자료	
통계/수치/자료/보고서	종합주가지수/보도자료/고객만족도/2007OECD보고서
11. 연구방법	
	조사분석/가치평가/동향예측/발전전망/저작비평/주제탐구

정보의 검색과 해당 용어의 의미구조 및 대상 주제 도메인을 구성하는 개념의 이해에 분명하게 도움이 된다. 그러나 주제 색인어를 통한 검색 결과의 적합성은 해당 주제 색인어의 포함여부가 아니라 해당 연구학술정보에서 그 주제 색인어가 갖는 기능과 역할에 의해 결정된다. 그러므로 연구학술정보 각각 도출된 주제 색인어 리스트는 해당 정보를 구성하는 주요 개념을 표현하는 용어들의 집합이 아니라 그 역할과 의미관계를 통해 해당 연구학술정보의 지식구조를 반영하는 구조화되어야 하며 연구자는 연구학술정보 각각의 지식구조에 대한 이해를 통해 연구프로세스 단계별로 요구되는 분석적 정보활동을 효율적으로 수행할 수 있다.

4. 온톨로지 기반 주제 색인어 구조화 방안

연구자가 주제 색인어를 통해 자신의 수요에 적합한 연구학술정보를 효율적으로 찾아 쉽게 접근하고 이해할 수 있으려면, 먼저 주제 색인어의 개념 이해를 통해 적합한 용어를 선정하고 검색할 수 있어야 한다. 주제명 표목(subject headings)은 대상 연구·학술정보가 포함하는 중요한 개념을 표현하는 색인어로서 해당 연구·학술정보에 대한 주제에 대하여 포괄적인 이해기반을 제시하는데 해당 개념을 포함하고 있는 정보에 대한 접근 경로

를 제공하고, 연구자는 최종적으로 그 색인어가 표현하는 개념 또는 대상이 그 해당 연구·학술정보에서 어떻게 다루어졌느냐에 따라 적합성 여부를 판단하게 된다. 연구자는 주제 색인어로 검색된 각각의 연구·학술문헌을 구성하는 개념들과 그 개념들의 의미관계 파악을 통해 해당 문헌의 지식구조를 이해하고 적합성을 판단하여 자신의 관점과 목적에 따라 전체 또는 일부의 지식을 참조하거나 활용하게 된다. 그러므로 연구자가 원하는 지식구조를 가진 연구·학술정보에 효율적 접근을 하기 위해서는 어휘적 지식구조와 함께 연구·학술문헌의 지식구조 및 이해형성 과정 관점에서 해당 문서의 지식구조를 반영하는 개념 체계에 대한 이해가 필요하다.

연구·학술정보의 개념체계는 담화 이론(dis-course theory) 및 텍스트 언어학(text linguistics) 기반의 연구·학술문헌의 개념구조 분석(Kirsz, J. 1991; .G. Kircz and F.A.P. Harmsze 2000; Waard, A. de, Breure, L., Kircz, J.G., & Oostendorp, H. van 2006), 연구수행프로세스 기반 개념구조 연구(Oddy, R. et al. 1992) 연구·학술문헌 개념구조 기반의 정보검색(Kando, N. 1997), 연구·학술정보의 수요 적합성 기준 및 정보행위 이해(Schamber, L. 1994), 인지적 관점에서 연구수행 단계별 연구·학술정보 이용자의 수요 연구(Wang, P. 1997) 등 다양한 관점에서 연구·학술정보의 효율적 의사소통 및 검색을 지원하는 지식기반으로서 다루어져 왔다. 정

보검색도구로서 주제 색인어 목록은 개별 연구 학술정보의 개념체계와 관련 용어들의 집합으로서 개별 용어 단위가 아니라 해당 연구 학술정보에서 갖는 상호 의미관계를 표현하는 지식구조로서 정의되어야 한다. 본 연구에서는 주제 영역의 지식체계를 반영하는 연구 · 학술문헌의 지식구조를 인문사회과학적 연구 수행을 위한 의미 및 이해형성 관점에서 정형화하고 이를 반영하는 온톨로지 기반 구조화된 시소러스 구축방안을 제시하고자 한다.

인문사회과학 분야 연구자는 일반적으로 사회 전반에서 발생하는 문제점을 ①다양한 현상과 이슈를 통해 파악하고 ②문제를 바라보는 시각 및 관점을 수립하며 ③관점을 뒷받침하는 이해와 주장의 확보 ④이해를 구성하는 개념과 정의를 통해 연구자는 이해형성의 기반을 구성한다. 이를 위해 연구자는 사회를 구성하는 ⑤기본 구성요소를 대상으로, 구성 요소간의 다양한 상호작용으로 나타나는 현상과 이슈의 규명을 위해 ⑥속성 파악을 통해 대상 객체를 구체적으로 식별하고 ⑦그 상호작용을 규정 · 통제 · 구성하는 체계 ⑧상호작용이 발생하는 시 · 공간 및 조건을 식별하고, ⑨대상의 특성 또는 상호작용의 효과를 관찰 · 인지 · 측정하기 위해 ⑩수집한 다양한 자료를 ⑪적합한 수단과 도구를 이용하여 ⑫다양한 방법으로 측정 · 파악하고 ⑬그 측정 결과, 즉 데이터 분석 또는 해석의 결과를 제시함으로써 연구결과의 타당성을 주장하는 다양한 성격과 유형의 연구를 수행한다. 연구 · 학술정

보는 이러한 학술적 의미형성 과정을 구성하는 개념들이 논리적 관계로 구성된 문서 단위 지식체계이다.

학술적 이해형성 과정에서 요구되는 개념 유형을 학술연구정보로부터 추출 또는 선정된 주제 색인어의 범주화에 적용함으로써 대상문서의 지식체계를 구성하는 개념의 범주화를 할 수 있다. 위의 열거된 인문사회과학 분야 연구자들의 일반적인 연구수행단계와 관련된 정보활동과 관련된 총 10개의 개념들을 용어로 표현하면 각각 개념 · 정의, 이해, 관점, 구성요소, 체계, 문제 · 이슈, 특성, 수단 · 도구, 데이터 · 자료, 연구유형 등 정의될 수 있다. 학술적 이해형성 관점의 연구학술정보의 개념체계와 각 주제영역별 도메인 지식체계의 개념체계는 각기 다른 관점의 독립적 지식구조를 갖는다. 예를 들어 위의 정의한 개념범주와 KRF의 인문사회과학 분야 주제 색인어 기반 용어 레지스트리(최석두 2007; 김이점, 이재운 2007)의 개념범주와 의미적 상관관계를 살펴보면 다음과 같다. KRF 용어 레지스트리의 13개 용어 클래스는 각각 인물, 조직기관, 재화, 자연, 각종 콘텐츠 등은 사회적 상호작용의 구성요소'로서, 법 · 제도 · 교육 · 의식 · 종교 · 문화" 등 사고와 행동양식의 형성에 영향을 주는 객체를 의미하는 개념은 체계와 대응할 수 있다. 주제어에서 범주화되지 않은 용어 중 현상에 대한 체계적인 이해'를 표현하는 용어, 즉 이론 · 원리 · 법칙 · 모델 개념을 표현하는 용어는 이해' 범주로, 그 이해를 논

리적으로 뒷받침하는 개념 또는 정의를 나타내는 용어를 개념·정의 범주와, 학제·학파·주의 등 현상을 바라보는 특정한 시각과 관점을 표현하는 용어는 관점 등으로 매핑될 수 있지만, 원칙적으로 도메인 지식체계를 구성하는 개념은 각각 독립적인 분류체계(domain taxonomy)에 의해 관계구조를 갖는다. 따라서 이론, 원리나, 개념이나 정의에 해당하는 용어들이 연구 주제가 될 수 있으며 각 개념범주에 속한 인스턴스들 간의 의미관계구조는 용어의 개념속성 및 구조에 따라 논리적으로 설정되는 것이 아니라 연구자의 관점과 주장을 반영하는 해당 연구학술정보에 한하여 유효한 범주와 관계에 의해 결정되는 고유의 지식이다.

〈표 8〉은 연구수행 및 학술적 이해형성 단계별로 요구되는 정보의 유형과 관련된 개념에 따라 정의된 주제어의 10개의 기본 개념범주에 해당하는 인문사회과학 분야 연구·학술

문헌의 개념체계와 연구·학술 어휘 관계구조와 그러한 범례적 연관 관계를 나타내는 것으로 이해함이 타당하다. 여기서 제시하는 개념 범주는 최소 온톨로지 구축 규칙(minimal ontology commitment)(Gruber, T. 1995)에 따라 정의된 일종의 프로토 타입으로서 상세한 연구자의 관점과 주장은 궁극적으로 대상 연구·학술정보를 통해 파악되어야 하나 연구자의 분석적 정보탐색에 필요한 수준의 주제 색인어의 구조화를 의미한다. 학술적 이해형성 관련 정보활동 세분화 수준에 따라 개념 범주의 추가적 정의나 주제 도메인별로 적절한 하위 범주의 점진적 세분화가 가능하다 (Shimpan, F.M and McCall, R. 1994).

연구자들은 연구를 통해 식별된 각 개념범주의 인스턴스들을 정형화된 논리적 연관관계를 통해 구체화하고 일정한 형식의 문서작성을 통해 연구 성과를 제시한다. 주제 색인어 리스트를 통한 연구·학술정보 지식구조 표현

〈표 8〉 연구·학술정보 개념구조 분석 사례(1)

개념범주	주제 색인어
문제/이슈	검색 효율성 평가
관점/주제분야	문헌정보학-정보검색
이해/모델	FRBR모형, OPAC모형
개념/정의	검색 효율성
구성요소/연구대상	(FRBR모형 및 OPAC모형 기반) 서지검색시스템
활용자료	음악자료
방법	구현형 질문, 표현형 질문
특성/측정대상 속성	재현율, 정확율, 검색시간
수단/도구	중다변 분석과 T-검증
근거 데이터	통계분석 비교
연구방법	비교 분석

을 위해 해당 연구 · 학술논문의 내용을 위에 제시된 연구 · 학술정보의 지식구조 및 연구프로세스의 정보활동 관점에서 기술할 수 있다. 예를 들어 FRBR 모형 기반 서지검색시스템의 검색효율성 평가 연구'라는 학술논문의 경우를 살펴보면, 해당 정보는 FRBR 모형 기반 서지검색시스템 검색효율성 평가' 수행을 위해 ①문헌정보학적 관점에서 ②FRBR 모형, OPAC 모형, 검색 효율성에 대한 이해에 기반하여 ③각 모형이 적용된 검색 시스템을 대상으로 ④음악자료에 대하여, ⑤구현형 질문, 표현형 질문을 이용하여 ⑥재현율, 정확율, 검색시간을 측정하고 ⑦중다변 분석과 T-검증을 이용한 통계분석 비교를 통한 연구 성과를 제시하고 있다. 해당 정보의 지식구조를 개념범주와 속성 값으로 표현하면 <표 9>와 같다.

개념 간의 관계를 용어 관계구조로 표현하면, 연구 · 학술정보의 핵심주제로서 문제, 즉 검색 효율성 평가'를 중심으로 각각의 주제 색

인어가 연구 · 학술정보의 논리체계에 따라 다음과 같은 의미적 관계를 형성한다.

문제: 검색 효율성 평가 - 관점: 문헌정보학

문제: 검색 효율성 평가 - 대상: FRBR 기반 서지검색시스템

문제: 검색 효율성 평가 - 자료: 음악자료

문제: 검색 효율성 평가 - 조사방법: 질의어 테스트

문제: 검색 효율성 평가 - 수단: 구현형 질문

문제: 검색 효율성 평가 - 수단: 표현형 질문

문제: 검색 효율성 평가 - 특성: 재현율

문제: 검색 효율성 평가 - 특성: 정확율

문제: 검색 효율성 평가 - 특성: 검색시간

문제: 검색 효율성 평가 - 해석방법: 통계분석 비교

문제: 검색 효율성 평가 - 도구: 중다변분석

문제: 검색 효율성 평가 - 도구: T-검증

문제: 검색 효율성 평가 - 연구방법: 비교 분석

실제 해당 학술논문의 저자 키워드는 FRBR

<표 9> 연구 · 학술정보 개념구조 분석 사례(2)

개념범주	주제 색인어
문제/이슈	온톨로지 구축 방안 연구
관점/주제분야	문헌정보학-지식기반
이해/모델	SKOS
개념/정의	온톨로지, (시소러스)
구성요소/대상	시소러스
활용 자료	한국역사 용어시소러스
특성/측정대상 속성	의미관계의 일관성
수단/도구	단계적 변환 알고리즘
근거 데이터	RDF문서
연구방법	용용 연구

모형, FRBR화, 검색 효율성, 서지검색시스템, 재현율, 정확률, 검색시간, 음악자료, 구현형 질문, 표현형 질문, FRBR Model, OPAC, KORMARC, FRBRizing, Retrieval Efficiency Recall, Precision, Retrieval Time로서 용어차원에서 해당 학술논문의 개념구조를 반영하고 있음을 알 수 있다.

또 다른 예로서 시소러스의 SKOS변환을 중심'으로 라는 부제를 갖는 시소러스를 활용한 온톨로지 구축방안 연구'라는 연구학술논문의 지식구조를 연구수행 프로세스 관점에서 기술하면 다음과 같다. 해당 정보는 시소러스를 활용한 온톨로지 구축방안 연구' 수행을 위해 ①문헌정보학적 관점에서 ②온톨로지 개념을 바탕으로 하는 ③SKOS(모델)에 기반하여 ④한국역사시소러스의 활용 및 ⑤단계별 알고리즘을 적용을 통해 ⑥변환된 RDF문서에 대해 ⑦의미관계 일관성을 측정된 결과를 연구성과로 제시하고 있으며 이를 나타내는 의미기반의 색인 <표 9>에서 구조화된 주제색인어 집합으로 표현될 수 있다.

해당 논문의 실제 주제 색인어는 시소러스, SKOS 온톨로지, 의미관계, 도메인, thesaurus, SKOS, ontology, semantic relation, domain로서 위의 FRBR 연구와 마찬가지로 수단/도구 개념에 해당하는 주제 색인어가 생략되어 있는데, 특별히 강조가 필요한 경우를 제외하고는 수단과 도구는 주제색인어 리스트에서 생략되는 경향이 있다. 아래는 해당 학술논문의 지식구조를 나타내는 온톨로지 기반

시소러스의 용어관계구조이다.

문제: 온톨로지 구축방안 연구 - 관점: 문헌정보학

문제: 온톨로지 구축방안 연구 - 이해: SKOS

문제: 온톨로지 구축방안 연구 - 개념: 온톨로지

문제: 온톨로지 구축방안 연구 - 대상: 시소러스

문제: 온톨로지 구축방안 연구 - 자료: 한국 역사용어 시소러스

문제: 온톨로지 구축방안 연구 - 특성: 의미 관계의 일관성

문제: 온톨로지 구축방안 연구 - 방법: 자동 변환 및 결과 보정

문제: 온톨로지 구축방안 연구 - 수단: 단계 별 변환 알고리즘

문제: 온톨로지 구축방안 연구 - 데이터: RDF문서

문제: 온톨로지 구축방안 연구 - 연구방법: 응용연구

연구·학술정보 중, 저널논문의 경우 위와 같은 의미적으로 정형화된 개념구조를 갖는데 대상 정보의 학술적 이해형성 관점에서 각 개념들의 유기적인 관계를 다음과 같이 기술할 수 있다.

연구·학술논문 J(journal article)는, 문제 X에 대한 해결방안으로서, P(perspective)의 관점에서, 개념 C(concept) 또는 정의 D(definition)에 기반하는, 이해 U(understanding)를 바탕으로, 대상 O(object)의 인스턴스로서 자료 M(materials)을 활용하여, 속성 P(property)를 도구 I(instrument)를 방법

M(Methods)을 통해 측정하여, 결과 데이터 R (result)의 해석과 분석결과를 제시하는 타입 T(type) 연구이다.

각각의 개념범주는 다음과 같은 논리적 의미관계로 연결되어 각각의 연구학술정보의 지식구조를 정형화된 형식으로 표현할 수 있다.

Understanding <isBasedOn> Concept, Definition

Understanding <hasDomain> Perspective

Understanding <isAppliedTo> Problem

Objects <isRelatedTo> Problem

Objects <hasProperty> Property

Materials <isInstanceOf> Objects

Property <isMeasuredBy> Methods

Methods <uses> Instrument, Tool

Data <isResultedFrom> Methods

Data <isAnalyzedBy> Methods or Instrument, Tool

연구학술정보는 위와 같은 형식의 개념 간의 논리구조로 개념 간의 관계를 표현하는데, 문학, 예술 등 각기 다른 논리체계를 갖는 주제 분야에 따라 다른 최적화된 형식으로 표현될 수 있다. 중요한 것은 용어단위의 시소러스와 달리 연구자는 문서단위의 지식체계를 반영하는 구조화된 의미 색인 온톨로지를 통해 관련 자료에 대한 주제적 접근뿐만 아니라 각 개념들이 해당 연구학술정보마다 어떻게 다루어졌는지 개념간의 학술적 관계구조를 파악할

수 있다는 점이다. 용어단위의 관계구조에서 표현되지 않거나 정형화할 수 없는 연구주제와 관련된 다양한 관련 개념, 정의, 이론, 수단, 도구, 자료, 데이터 등 사이에 성립하는 다양한 논리적 연관관계를 명확하고 구체적으로 명시함으로써, 연구자는 개념속성과 제약 조건 등을 이용하여 대상 문서정보들로부터 적합한 지식구조를 갖는 정보를 체계적이고 효율적으로 탐색할 수 있다.

한편 연구 · 학술정보의 지식구조를 구성하고 있는 각각의 개념범주들은 해당 주제영역에서 유효한 도메인 분류체계를 가질 수 있는데 그 조합의 패턴을 분석하여 해석 연구, 탐구 연구, 비교 연구, 조사 연구, 장기 연구, 사례연구, 메타분석 연구, 다학제 연구 등의 연구유형을 규정하는 공리적 지식의 정의가 가능하다(Ziemski 1975). 또한 유사한 연구주제에 대하여 각 개념범주별로 관점, 개념, 대상, 방법, 수단, 도구, 결과데이터, 분석 및 해석 등 각 조합에 따른 여러 가지 추론 규칙의 도출도 가능하다. 예를 들어, 각 연구 · 학술정보를 구성하는 개별 개념 간의 성립하는 의미 관계 분석을 통해 각 정보 간의 성립하는 수사적 관계에 기반한 담화지향 온톨로지를 설계할 수 있다.

유사 동일한 연구주제 X를 다루는 연구학술논문 J1과 J2의 개념 C1, C2가 대립적 관계이면 두 정보 간에는 학술적으로 의미적 불일치 관계가 성립한다.

⇒ J1 <isInconsistentWith> J2

유사 동일한 연구주제 X를 다루는 연구학술논문 J1과 J2의 이해 U1, U2가 USE/UF 또는 BT/NT 관계이고 방법 M1과 M2가 다른 대안적 선택 관계가 성립한다.

⇒ J1 <isAlternativeFor> J2

연구주제 X를 다루는 연구학술정보 J1, J2에 대하여 각각의 개념 C1, C2가 USE/UF 또는 BT/NT 관계이고 이해 U1, U2가 다른 경쟁적 관계가 성립한다.

⇒ J1 <isCompetitiveTo> J2

각기 다른 관점 P1, P2의 연구학술논문 J1, J2가 각각의 개념 C1, C2, 이해 U1, U2, 방법 M1, M2가 유사·동일하면 상호 응용 또는 적용 관계가 성립한다.

⇒ J1 <supports> J2 또는 J2 <supports> J1

연구·학술정보 간의 담화적 관계의 정의 및 분석을 통해 해당 주제영역의 학술 커뮤니케이션 의미 네트워크를 구성할 수 있다. 연구자는 이를 통해 대상 주제영역의 지식 구조를 보다 심층적으로 이해 할 수 있고, 연구단계별로 필요한 정보를 키워드 검색과 더불어 상호 의미 관계를 통해 상대적 적합성을 판단할 수 있어 다양한 정보수요와 목적에 맞는 적합한 질의 방법을 통해 정보의 효율적 검색과 정확한 접근이 가능하다.

5. 결론

본 연구에서는 연구·학술정보의 효과적인 검색과 접근을 위해 대상 정보의 지식구조를 반영하고 정보 간의 의미관계 추론을 지원하는 온톨로지 기반의 주제 색인어 시소러스 구축방안을 제시하였다. 색인어 기반 검색 효율성 제고를 위해 용어의 관계의 다의성과 모호성 문제해결 및 확장검색지원 등 용어관계 구조화 및 세분화를 중심으로 수행되어온 기존의 시소러스 연구의 문제와 한계 해결을 위해서 검색 대상 정보의 지식구조를 반영하는 지식조직체계가 뒷받침되어야 함을 확인할 수 있었다. 주제 도메인 구조 및 상세화 수준의 설정을 위한 기준으로서 연구수행 프로세스와 그 과정에서 발생하는 연구자의 정보수요 유형을 분석하고 인문사회과학 분야 연구학술논문의 지식구조를 표현하는 기본적 개념범주와 유기적 관계 정의를 통해 문서 단위의 의미 색인을 지원하는 연구·학술정보 온톨로지를 정의하였다. 정보의 효율적인 탐색과 적합성 파악을 지원하는 지식자원의 구축을 위해 특정 지식계층의 개선이나 확장보다는 대상 정보의 지식체계를 개념적으로 각각 용어, 문서, 담화구조의 3개 하위계층으로 구분하고 상호 유기적 연계방안에 초점을 맞추어 지식조직체계 구성에 방안을 모색하였다. 그 결과 주제 색인

어 리스트를 대상정보를 구성하는 주요 개념을 표현하는 용어의 집합이 아닌 문서를 구성하는 개념과 개념 간의 유기적 관계 정의를 통해 해당 정보의 지식구조를 반영할 수 있도록 구조화된 의미기반 주제 색인 방법을 제시하였다. 이를 기반으로 유사어·반의어 등 기존 시소러스 용어관계구조의 관계 속성과 문서의 개념구조 및 개념 간 논리적 관계 속성, 그리고 연구·학술정보 간에 성립하는 담화적 의미관계 분석을 지원하는 추론규칙 정의 등 각각의 계층의 정보가 상호 유기적 관계를 갖는 지식조직체계로서 구조화된 학술커뮤니케이션 네트워크의 구성과 새로운 의미기반의 정보탐색 및 분석 서비스를 지원하는 지식자원으로서 활용방안을 살펴보았다. 새로운 의미기반의 정보탐색 및 분석 서비스의 실제 구현을 위해 고려되어야 할 과제는 다음과 같이 정리할 수 있다.

첫째, 문서의 지식체계를 표현하는 개념 구조화 수준은 연구자 정보 활용 상황에 대한 이해를 바탕으로 결정되어야 한다. 예를 들어 연구 프로세스 단계별로 요구되거나 수행되는 정보활동에 대한 좀 더 체계적인 분석을 통해 복합적인 정보수요를 효과적으로 만족시킬 수 있는 공리적 지식에 대한 정의를 추가 보완할 수 있다.

둘째, 연구자의 관점 및 목적별로 변화하는 대상 정보의 적합성 관련 의사결정의 지원을

위해 정보 간 성립하는 의미관계 유형정의와 세분화가 이루어져야 한다. 해당 주제 도메인의 연구방법론과 연구형식을 기준으로 연구학술정보의 개념구조를 정형화하는 방법도 유용한 접근 방식으로서 연구유형 분류체계의 정의를 고려할 필요가 있다.

셋째, 효율적 지식자원 구축·관리체계의 수립이다. 초기 기반 마련 및 안정화를 위해 특정 주체가 분산된 정보자원을 수집하여 가공하는 정보처리 절차가 불가피 할 수 있으나 중장기적으로는 많은 비용과 노력이 소요됨으로 지속적 수행이 불가능하다. 따라서 기존의 저자 주제 색인어 리스트와 같이 개별 단위정보별로 대상 문서의 지식구조를 표현하는 온톨로지 기반의 의미 색인 정보를 저자 또는 정보 전문가가 기술할 수 있도록 표준화된 입력 환경을 제공하고 공유함으로써 정보수요자가 제공자로서 참여하는 커뮤니티 형태가 바람직하며 정보의 작성 및 배포단계에서 필수적으로 작성·제공되는 메타데이터로서 표준화가 요구된다.

참고문헌

- 고영만. 2006. 시소러스 기반 온톨로지에 관한 연구. 『정보관리』, 5: 5-22.
- 김원중, 양재동. 2007. 시각화된 환경에서 다차원 관점을 지원하는 객체기반 패시 시소러스 관리 시스템 모델의 정형화 및 구현. 『한국정보과학회논문지 : 소프트웨어 및

- 응용』, 34(9): 828-847.
- 김이겸, 이재윤. 2007. 『용어 데이터레지스트리 설계 및 구축지침』. 서울: 한국학술진흥재단.
- 김지훈, 김태수. 2006. 용어정의와 관계추출을 통한 시소러스 확장에 관한 연구. 『한국문헌정보학회지』, 40(1): 293-314.
- 김태수. 2001. 용어정의를 도입한 시소러스 개발 연구. 『정보관리학회지』, 18(2): 231-254.
- 유영준. 2005. 온톨로지의 개념간 관계 설정을 위한 AGROVOC 시소러스의 분석에 관한 연구. 『정보관리학회지』, 22(1): 125-144.
- 이정연, 김정민, 최석두, 김이겸. 2007. 기초학문 자료 메타데이터 설계 분석 및 온톨로지 적용 방안 연구. 『한국문헌정보학회지』, 41(2): 291-316.
- 정한민, 성원경, 박동인, 황순희. 2005. 과학기술 분야 시소러스상에서의 패킷들. 『한국정보과학회 가을학술발표문집(II)』, 32(2): 535-537.
- 조현양, 남영준. 2004. 시소러스와 온톨로지의 상호 호환성에 관한 연구. 『정보관리학회지』, 21(4): 27-47.
- 최석두, 이우범, 김이겸, 최상기, 한상길, 이정연. 2006. 대단위 우리말 온톨로지 구축을 위한 시소러스의 개발. 『정보관리학회지』, 12: 147-164.
- 최석두. 2007. 『인문사회과학분야 용어데이터 레지스트리 설계 보완 및 활용연구』. [서울]: 한국학술진흥재단. 정책연구-2007-025-지식확산.
- 한성국, 이현실. 2004. 시소러스를 활용한 온톨로지 구축방안 연구. 『한국비블리아학회지』, 17(1): 285-303.
- 한상길. 1999. 『시소러스 용어관계의 확장에 관한 연구』. 박사학위논문, 중앙대학교, 문헌정보학과.
- 황순희, 윤애선. 2005. 의미적 준거의 세목화를 고려한 과학기술분야 시소러스 구축. 『한국어 의미학』, 18: 99-124.
- Aitchison, J and Gilchrist, A., Bawden, D. 2000. Thesaurus Construction and Use: A Practical Manual, 4th ed. Europa publication.
- Bargmeyer, B. 2005. "Establishing Ties Between Metadata Registries and Terminology." 8th Open forum on Metadata Registries.
- Beissel-Durrant, G. 2004. "A typology of research methods within the Social Sciences". NCRM Working Paper, Southampton, UK, University of Southampton.
- Broughton, V. 2006. "The need for a faceted classification as the basis of all methods of information retrieval." Aslib Proceedings: New Information Perspectives, 58(1/2). DOI 10.1108/

- 00012530610648671.
- Buckingham S., Motta, E., and Dominique J. 1999. "Representing Scholarly Claims in Internet Digital Libraries: A Knowledge Modelling Approach." Proc. 3rd. European Conference on Research and Advanced Technology for Digital Libraries, Paris, 22-24, Sept. 1999.
- Buckingham, S. and Motta, E., Dominique, J. 2000. "Scholonto: An Ontology-Based Digital Library Server for Research Documents and Discourse." International Journal on Digital Library, 3: 237-248.
- Buckingham S. and Selvin A.M. 2000. "Structuring Discourse for Collective Interpretation." Proc. Conference on Collective Cognition and Memory Practice: Distributed Collective Practices 2000, Paris, 19-20, Sept. 2000.
- Buckingham, S. 2006. "Sensemaking on the Pragmatic Web: A Hypermedia Discourse Perspective." Proc. PragWeb '06: 1st International Conference on the Pragmatic Web, Stuttgart, 21-23 Sept. Springer-Verlag: Berlin, 2006.
- Ceusters, W. and Smith B., Goldberg L. 2005. "Terminological and Ontological Analysis of the NCI Thesaurus." Methodes of Information in Medicine, 44: 498-507.
- Fischer, D. 1998. "From Tesauri towards Ontologies?" Proc. of 5th Int. ISKO-Conference, 18-30, Aug. Lille.
- Gruber, T. 1995. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing." International Journal of Human-Computer Studies, 43: 907-928.
- Hunter, J. 2001. "MataNet - A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains." Journal of Digital Information, 1(8).
- Kando, N. 1997. "Text-level structure of research papers: Implications for text-based information processing systems." In Proceedings of the British Computer Society Annual Colloquium of Information Retrieval Research, Oct. 1997 68-81.
 <<http://www.rd.nacsis.ac.jp/~kando/papers/bcs97-textstru.ps>>.
- Kirsz, J. 1991. "Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval." Journal of Documentation, 47(4): 354-372.
- Kircz, J and F.A.P. Harmsze. Modular

- scenarios in the electronic age. Conferentie Informatiewetenschap 2000. Doelen, Rotterdam 5 april 2000. In: P. van der Vet en P. de Bra (eds.) CS-Report 00-20, Proceedings Conferentie Informatiewetenschap 2000, De Doelen Utrecht (sic), 5 april 2000. pp. 31-43.
- Liang, A., Lauser, B., Sini, M., Keizer, J., Katz, S. 2006. "From AGROVOC to the agricultural ontology service/concept server: An OWL model for managing ontologies in the agricultural domain." Atlanta, in Proceedings of OWL: Experiences and Directions Workshop.
- Maple, A. 1993 "Faceted Access: A Review of Literature." Feb. 1995
 <http://www.music.indiana.edu/tech_s/mla/facacc.rev>.
- Matthews, B. 2004. "CRISs, Thesauri and Semantic Web." Proc. of the 7th International Conference on Current Research Information Systems, eds. Annemarie Nase, Geert Van Grootel, 113-124.
- Oddy, R., Liddy, E., Balakrishnan, B., Bishop, A., Elewononi, J., Martin, E. 1992. "Towards the use of situational information in information retrieval." *Journal of Documentation*, 48(2): 123-171.
- Sager, J.C. 1990. *A Practical Course in Terminology Processing*. Amsterdam, John Benjamins.
- Sager, J.C. and L'Homme M.C. 1994. "A Model for the Definition of Concepts: Rules for Analytical Definition in Terminological Databases." *Terminology*, 1(2): 351-373.
- Sager, J. C. and Ndi-Kimbi A. 1995. "The Conceptual Structure of Terminological Definitions and their Realizations." *Terminology*, 2(1): 61-81.
- Schamber, L. 1994. "Relevance and Information Behavior." *Annual Review of Information Science and Technology*, 29: 3-48.
- Shimpan, F.M and McCall, R. 1994. "Supporting Knowledge-base Evolution with Incremental Formalization." in Proc. ACM CHI '94 285-291, ACM Press.
- Soergel, D. 2004. "Building a rich ontology from AGROVOC." *FAO Agricultural Ontology Server Workshop*, Beijing, Apr. 27-29.
- Tudhope, D., and Binding C. "Faceted Thesauri." *Axiomathes* 18: 211-222, DOI 10.1007/s10516-0080-9031-6.

- Waard, A. de, Breure, L., Kircz, J.G., & Oostendorp, H. van, 2006. Modeling Rhetoric in Scientific Publications In Vicente P. Guerrero Bote (Editor) (Ed.), Current Research in Information Sciences and Technologies (pp. 352-356). International Conference on Multidisciplinary Information Sciences and Technologies InSciT2006 Badajoz, Spain: 25-28
- October 2006.
- Wang, P. 1997. "Users' information needs at different stages of a research project: a cognitive view." Proceedings of an international conference on Information seeking in context, 307-318, Tampere, Finland.
- Ziemski, S. 1975. "Typology of Scientific Research." Journal of General Philosophy of Science, 6(2): 276-291.