

한국학술지인용색인(KCI)의 인문학, 사회과학, 예술체육 분야 저자키워드의 의미적 관계 유형 최적화 연구*

A Study on the Optimization of Semantic Relation of Author Keywords in Humanities, Social Sciences, and Art and Sport of the Korea Citation Index (KCI)

고 영 만 (Young Man Ko)**

송 민 선 (Min-Sun Song)***

이 승 준 (Seung-Jun Lee)****

목 차

- | | |
|--------------------------|---------------|
| 1. 서 론 | 4. 관계 유형의 최적화 |
| 2. STNet의 용어 간 관계 유형 | 5. 결 론 |
| 3. 관계 유형의 최적화를 위한 데이터 분석 | |

초 록

본 연구의 목적은 KCI의 인문, 사회, 예술체육 분야 저자키워드를 기반으로 구축한 구조적 학술용어사전 데이터베이스 시스템 STNet에 실제 입력된 용어들의 의미적 관계들을 분석하여 관계 유형을 최적화하고 관계명을 세분화하는 방안을 기술하는 것이다. 분석 결과, 관계 유형과 관계명을 최적화하기 위해서는 빈도수에 의한 관계명 신설 또는 관계의 유형화 제한, 관계의 방향성 고려, 기존 관계명 반영의 4가지 논리적 기준이 필요한 것으로 나타났다. STNet의 "실존인물" 클래스를 테스트베드로 삼아 이 기준을 적용해 관계 유형을 최적화하고 관계명을 세분화한 결과 원래 RT, RT_X, RT_Y로 연결된 1,743건의 관계 중 1,135건이 세분화되는 것으로 나타났다. 이는 약 65%의 RT 관계가 구체화된 것으로 사전 구축 및 검색 단계에서의 유용성 측면에서 상당한 의미가 있음을 보여주는 것이다.

ABSTRACT

The purpose of this study is to analyse the semantic relations of terms in STNet, a structured terminology dictionary based on author keywords of humanities, social sciences, and art and sport in the Korea Citation Index (KCI) and to describe the procedure for optimizing the relation types and specifying the name of relationships. The results indicate that four logical criteria, such as creating new names for relationships or limitation of typing the relationship by the appearance frequency of same type, consideration of direction of relationship, reflection to accept the existing name of relationships, are required for the optimization of the typing and naming the relationships. We applied these criteria to the relationships in the class "real person" of STNet and the result shows that 1,135 out of 1,743 uncertain relationships such as RT, RT_X or RT_Y are specified and clarified. This rate of optimization with ca. 65% represents the usefulness of the criteria applicable to the cases of database construction and retrieval.

키워드: 저자키워드, 학술용어, 속성, 한국학술지인용색인, KCI, 구조적학술용어사전, STNet 의미적 연결관계, 관계 유형
Author Keyword, Scientific Term, Property, Korea Citation Index, KCI, Structural Terminology Net, STNet Semantic Relation, Relation Type

* 이 논문은 2012년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2012S1A5B4A01035131).

** 성균관대학교 문과대학 문헌정보학과 교수(ymko@skku.edu) (제1저자)

*** 성균관대학교 정보관리연구소 연구원(songser@skku.edu) (교신저자)

**** 성균관대학교 정보관리연구소 연구원(dino@skku.edu) (공동저자)

논문접수일자: 2015년 1월 26일 최초심사일자: 2015년 2월 6일 게재확정일자: 2015년 2월 10일
한국문헌정보학회지, 49(1): 45-67, 2015. [http://dx.doi.org/10.4275/KSLIS.2015.49.1.045]

1. 서론

1.1 연구 목적

학술 논문의 저자키워드는 논문의 저자들이 가장 핵심적이라고 판단해 추출한 학술 용어로 구성되어 있으며, 의미적인 면에서 해당 논문의 주요 내용들과 관련이 있다. 저자키워드는 또한 해당 학술 논문이 속한 학문 영역이나 주제 분야의 특성과 관련을 맺고 있으며, 논문과 관련된 기본적 개념이나 성질을 나타내는 내포적인 속성 및 해당 개념의 구체적 예시나 연관 관계를 나타내는 외연적 속성 등을 반영한다. 그리고 동일한 저자키워드를 공유하는 학술 논문들은 대체적으로 유사한 개념 영역에 속하게 된다.

저자키워드가 가지는 이러한 특성들은 연구자들이 학술 논문 간의 의미적 연관관계를 파악하는데 많은 도움을 줄 수 있다. 따라서 키워드 간 연결망을 구축하여 특정 학문 분야의 지식구조를 분석하거나(박옥남 2011; 장령령, 홍현진 2014), 학술 논문간의 상호 연관성을 밝히 고자 하는 연구들이 이루어져 왔다(양창진 2010; 이해영, 궤승진 2011). 그리고 저자키워드를 포함한 주요 학술용어를 중심으로 시소러스나 온톨로지를 개발하여 이를 의미적 연관 검색에 적용하고자 하는 연구도 꾸준히 이루어지고 있다(조현양, 남영준 2004; 유영준 2005; 고영만 2006; 이재호 외 2014).

기본적으로 시소러스나 온톨로지는 용어 간의 의미 관계를 구체화하여 정보 자원을 보다 효율적으로 검색하기 위한 목적에서 고안된 것이다. 따라서 연구자들이 저자키워드를 통해

학술 논문 간의 의미적 연관관계를 파악할 수 있는 시스템을 만들기 위해서는 저자키워드가 가지는 개념 간의 관계를 분석해서 다양한 연관관계 유형을 밝혀내는 작업이 선행되어야 한다. 그리고 이러한 작업을 통해 의미적 연관관계를 구체적으로 잘 정의하고, 정의된 관계를 바탕으로 추론 규칙을 용이하게 생성시킬 수 있어야 한다.

백지원과 정연경(2005)은 시소러스나 온톨로지의 용어 관계에 관한 기존의 연구 사례 29건을 분석하여 동등관계, 계층관계, 연관관계로 분류한 관계의 세부 유형 총 720여개를 정리하였다. 그러나 각 관계의 구분 기준이 불분명한 경우가 많아 용어 관계 및 지식 조직 체계를 다루는 학문 분야들 간의 협력 작업과 용어 관계를 단일 체계로 조직화할 수 있는 방법이 필요한 것으로 나타났다. 이와 관련하여 백지원(2005)은 패킷 분석 방식을 적용하여 연관관계를 세분화하는 방법과 함께 용어 간 상호 관계의 논리적 특성을 고려한 역관계 설정 및 실제 용어 출현 양상을 고려한 관계 분류 방법을 제안하였다.

지금까지 이루어진 용어의 의미 관계 구축 연구를 살펴보면 주로 텍스트 마이닝 기술을 이용해 텍스트 내에 존재하는 동사 및 관련 용어들을 추출해 설계하는 방법을 사용하거나(임수연, 박성배, 이상조 2005; 정현기, 김유섭 2008; 황미영 외 2012), 기존에 구축된 유사한 프로젝트의 관계 구조를 매칭하는 자동화 방법론을 적용하고 있다(Maedche and Staab 2002; 공현장 외 2005; 한성국, 이현실 2006; Akbari and Fathian 2010). 물론 실제 적용 과정에서의 의미적 오류를 줄이기 위한 작업에서는 대부분 사람이 직접 관계 유형들을 검토하는 방법을

사용하기도 한다(Zhitomirsky-Geffet and Erez 2014). 또한 사람이 직접 관계명을 생성하거나 검토한 연구도 있으나 대체로 유사 연구들에서 추출한 관계 유형들을 정리·분석하는 정도에 머물고 있다(정현숙, 최병일 2005; 조이현 외 2006; 김규환 외 2008).

기존에 제시된 용어 및 개념 관계 유형 중 계층관계는 구체적으로 정의하는 것이 비교적 용이하며, 기계적 처리에 의한 자동화 방법론을 적용하는 것이 어느 정도 가능하다. 그렇지만 동의 관계, 유의 관계, 연관 관계를 구체화하는 문제는 적용 목적이나 개발자의 관점 등에 따라 논리적인 기준이 다를 수 있어 그 해결이 쉽지 않다(전말숙 1998). 따라서 정보 자원 검색에 있어서 최적화된 의미 관계를 도출하기 위해서는 궁극적으로 사람이 직접 관계의 의미를 판단하고 부여한 데이터를 토대로 유형화하는 작업이 필요하다.

이와 관련하여 본 연구는 저자키워드를 기반으로 구축한 구조적학술용어사전 데이터베이스 'STNet(<http://www.stnet.re.kr>)'을 대상으로, STNet 구축 전문가 인력들이 구축 과정에서 실제로 생성하고 연결한 용어 간 관계 유형을 분석함으로써, 관계 유형 세분화의 최적화 방법을 기술하는 것을 목적으로 한다. STNet은 상호연관성을 갖는 연구성과물에 대한 지식지도를 형성하기 위해 한국학술지인용색인(Korea Citation Index, 이하 KCI)의 인문학, 사회과학 및 예술체육 분야 저자키워드를 유사한 속성을 가진 클래스(개념범주)에 따라 분류하고, 그 클래스가 가지는 속성을 체계화 한 다음, 그 속성에 따라 해당 학술 용어 하나하나의 의미를 정의하고 또 타 학술 용어와의 의미 관계를 정의

한 데이터베이스 시스템이다. 따라서 본 연구는 전문 인력이 직접 판단하여 구축한 STNet을 테스트베드로 삼아 용어 간 의미적 관계 유형을 분석한 사례 연구이다.

1.2 연구 방법

STNet은 대상 도메인이 비교적 넓고, 해당 도메인을 구성하는 클래스 텍사노미와 관계 유형의 수가 많고 복잡하며, 이에 따라 실제 시스템이 포함하는 데이터 또한 매우 방대하다. STNet이 구축 대상으로 선정한 용어는 KCI의 인문, 사회, 예술체육 분야 학술논문 중 2007년에서 2012년 9월까지 발행된 245,409편의 논문에 수록된 저자키워드 약 1백만 개 중 출현빈도가 2 이상인 저자키워드 약 6만여 개이다. STNet은 선정 용어들의 의미와 관계를 정의하는 과정에서 파생되고 확장된 학술용어를 포함하고 있으며, 2014년 12월 16일 현재 121,798개의 용어군 내에 371,565건의 연결 관계가 형성되어 있다. 따라서 기존 연구에서 생성된 관계 유형들만으로 STNet 구축을 위해 선정한 6만여 저자키워드로부터 파생되고 형성된 관계들을 모두 수용하는 것은 불가능하다. 이러한 문제를 해결하고 용어 간의 관계 유형들을 보다 효율적으로 재구조화하기 위해 STNet은 데이터베이스 구축 초기 단계부터 '관계속성' 항목을 도입하였다.

'관계속성' 항목은 STNet 구축 인력들이 데이터 입력 과정에서 용어 간의 관계명을 부여할 때, 그 시점까지 생성된 관계명의 리스트에 정확하게 일치되는 관계명이 없다고 판단되는 경우 사용하는 항목이다. 적합한 관계명이 없

는 경우 STNet 구축 인력은 관계명 리스트에서 가장 유사하다고 판단되는 관계명을 일단 적용하고, 추천하는 관계 유형을 '관계속성' 항목에 텍스트 값 형식으로 제안하게 된다. 이렇게 제안된 관계 유형은 추후 검수자와의 협의를 거쳐 선택적으로 수용되거나 더 적절한 유형으로 수정된다. STNet의 '관계속성' 항목은 모든 용어 간의 관계명 적용에 사용할 수 있으며, 특히 시소러스의 연관어 지시기호 "RT(Related Term)" 관계를 구체화시키는데 유용하다. 구축자나 검수자 입장에서도 '관계속성' 항목을 사용함으로써, "X 용어와 Y 용어가 연관이 있기는 하지만, STNet에서 제공하는 관계명 중에서는 정확히 대입되는 것이 없는데도 불구하고 그 중 하나는 반드시 적용해야 한다"는 문제로부터 자유로울 수 있다.

용어 간 관계 유형의 세분화는 궁극적으로 관계 구분 판단에 있어서의 논리성 문제와 유용성의 문제를 해결하는 것으로 수렴된다. 논리성의 문제는 관계 유형을 정할 때 구축 목적에 따라 달라지는 판단의 기준 문제를 어떻게 논리적으로 해결할 것인가에 관한 것이며, 유용성의 문제는 관계 유형의 세분화가 실제 데이터베이스 구축과 검색 과정에서 입력자와 이용자에게 유용하게 활용될 수 있는가의 문제이다. 본 연구의 목적을 달성하기 위해서는 STNet의 구축 과정에서 실제로 생성하고 연결한 용어 간 관계 유형을 분석하는 것이 필요하다. 그러나 STNet의 연결관계 37만여 건을 모두 분석하는 것은 불가능할 뿐 아니라 연구의 효율성 측면에서도 적절하지 못하다. 따라서 본 연구에서는 STNet에서 가장 많은 41,420개의 연결관계를 가지고 있는 클래스 「실존인물」에 속하는 용어

를 대상으로 STNet의 구축 전문 인력들의 의견과 관련 데이터를 분석하고, 분석 결과를 토대로 관계 유형의 최적화를 위한 논리적 기준과 실제 활용 측면에서의 유용성을 검토함으로써 연구의 목적을 달성하고자 한다.

2. STNet의 용어 간 관계 유형

2.1 STNet의 메타데이터 항목

다른 유사한 시스템에 비해 STNet이 가지고 있는 차이점은 용어사전 데이터베이스 구축과정 전반에 석사과정 이상의 전문 인력이 직접 관여해 작업을 수행하며, 클래스 텍사노미가 상향(bottom-up)식으로 구축되었다는 것이다(고영만, 김비연, 민혜령 2014). STNet 구축 인력은 선정된 6만여 저자키워드 하나하나의 의미를 해당 클래스의 속성에 따라 정의하고, 저자키워드의 클래스 속성을 비롯해 다른 저자키워드 혹은 기타 연관되는 학술용어와의 관계를 유형화한다.

STNet의 구축 과정은 선정된 학술용어의 담당자별 배포, 클래스 속성 등의 메타데이터 항목 입력, 입력 완료 용어의 검수 요청, 검수자의 검수, 검수 완료의 단계로 이루어진다. 용어의 클래스 속성 항목 및 다른 용어와 맺게 되는 관계 유형은 구축자와 검수자가 관련 학술 논문을 비롯한 연관 정보들을 참조하고 협의해서 입력한다. 입력을 위한 메타데이터 항목들은 크게 용어의 의미, 클래스, 관계에 관한 것으로 구별된다(〈표 1〉 참조).

〈표 1〉 STNet의 메타데이터 항목

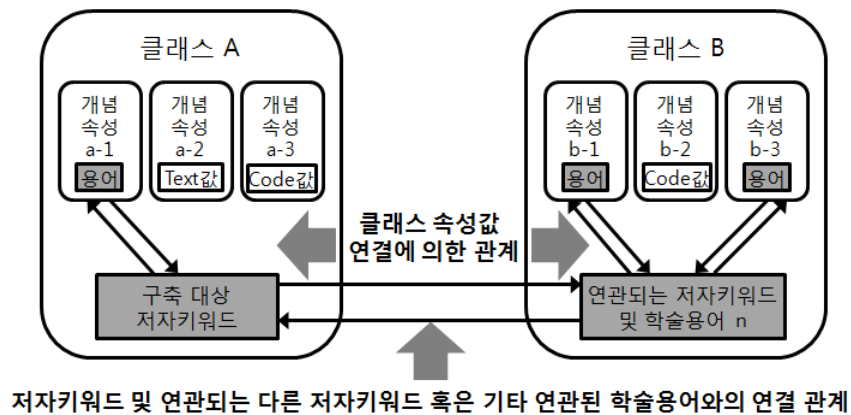
구분	항목	내용
용어 정보	연구분야	용어가 속한 논문 게재 학술지의 분야 (한국연구재단 연구분야분류표 기준)
	용어명	구축 대상 용어
	동의어	구축 대상 용어와 동의 관계인 용어 우선어-비우선어(USE-UF) 관계의 용어도 해당
	한자어	구축 대상 용어에 대응하는 한자어
	외국어	구축 대상 용어에 대응하는 외국어
	음차어	구축 대상 용어가 외국어인 경우 해당 음차어
클래스 정보	정의	구축 대상 용어에 대한 정의 (해당 학술 논문, 전문용어사전 또는 백과사전 활용)
	클래스명	STNet 클래스 텍사노미 기준
관계 정보	클래스 속성	항목값 유형: 객체(object), 텍스트(text), 코드(code) 필요에 따라 세분화된 클래스 속성 항목 추가 가능
	관계명	구축 대상 용어가 다른 용어와 갖는 관계
관계 속성	관계속성	관계 유형에 따라 세분화된 관계속성 항목 추가 가능

2.2 STNet의 용어간 의미적 관계 구성

STNet의 용어 간 의미적 관계는 크게 두 가지 종류로 나누어진다. 하나는 저자키워드와 그 저자키워드가 속한 클래스의 속성값이 연결되는 관계이며, 다른 하나는 저자키워드와 연관된

다른 저자키워드 또는 저자키워드와 기타 연관된 학술용어 간에 맺어지는 관계이다(〈그림 1〉 참조).

STNet 구축대상 용어인 저자키워드가 클래스의 속성값과 맺는 관계를 구체적으로 설명하자면 「예술작품명」이라는 클래스에 속하는 저



〈그림 1〉 STNet의 용어 간 의미적 관계 종류

자키워드와 클래스의 속성인 ‘저작자’, ‘저작년(시대)’ 등의 속성값들 사이에 형성되는 관계라 할 수 있다. 예를 들면 클래스 「예술작품명」에 속하는 저자키워드 ‘모나리자’가 속성값인 저작자 ‘레오나르도 다 빈치’, 저작년(시대) ‘16세기’와 형성되는 관계이다. STNet에서 클래스의 속성값은 객체(object) 값과 텍스트(text) 값, 코드(code) 값으로 구분되며, 저자키워드와 관계가 형성되는 것은 용어로 입력되는 객체값에 한한다. 2014년 12월 16일 현재 STNet에 저장된 객체값은 총 327종의 클래스 속성에 부여된 63,101건이다(〈부록 1〉 참조).

STNet에서 하나의 저자키워드는 해당 클레

스의 속성값과의 관계 외에도 다른 저자키워드 또는 다른 학술용어와 의미적 연결관계를 가질 수 있다. 이렇게 연결되는 STNet의 관계명은 시소러스에서 사용하는 계층관계, 동의관계, 연관관계와 기존의 전말숙(1998), 백지원(2005), 남영준(2009) 등의 연구에서 제시된 관계 유형을 참고하여 하향(top-down) 식으로 구축되었다. 2014년 12월 16일 현재, STNet에서 사용하고 있는 이러한 용어 단위 간 관계명은 역관계를 포함하여 총 100개이며(〈표 2〉 참조), 100개의 관계명을 통해 STNet 용어들 간에 형성된 실제 관계 연결 수는 308,464건이다(〈그림 2〉 참조).

〈표 2〉 STNet의 저자키워드 간 관계명(2014년 12월 16일 기준)

관계 유형	관계명	역(Inverse) 관계명	
계층관계	속-종	isA	
	부분 - 전체	BT	NT
		hasMember	isMemberOf
		containsSubstance	isSubstanceContainedIn
		hasComponent	isComponentOf
		hasIngredient	isIngredientOf
		spatiallyIncludes	isSpatiallyIncludedIn
	hasBranch	isBranchOf	
사례	hasInstance	isInstanceOf	
동의관계	LT	PT	
	Use	UF	
반의관계	hasSynonym		
	hasAntonym		
연관관계	RT_X	RT_Y	
	RT		
	hasIssue	isIssueIn	
	conceptuallyRelatedTo	isConceptOf	
	hasPhenomenon	isPhenomenonOf	
	hasForm(appearance)	isForm(appearance)Of	
	hasKind	isKindOf	
	affects	isAffectedBy	
hasDegree	isDegreeOf		

관계 유형	관계명	역(Inverse) 관계명	
연관관계	물리적	hasTributary	isTributaryOf
		isConnectedTo	
	기능적	apply	isAppliedBy
		hasMeasurement	isMeasurementOf
		hasEvaluation	isEvaluationOf
		analyze	isAnalyzedBy
		hasOpposition	isOppositionOf
		manage	isManagedBy
	공간적	traverses	isTraversedBy
		isAdjacentTo	
		surrounds	isSurroundsBy
	시간적	co-occurs-with	
		precedes	follows
	개념 속성 연관	hasPurpose	isPurposeOf
		hasTheory	hasAdvocate
		hasCreator	hasWork
		causes	isCausedBy
		hasSubject	isSubjectIn
		originateFrom	isOriginOf
		hasResult	isResultOf
		hasFounder	isFounderOf
		hasState	isStateOf
		hasSolution	isSolutionFor
		hasSupplement	isSupplementOf
		hasProperty	isPropertyOf
		hasAgent	hasPatient
		hasLocation	-
		produces	isProducedBy
	hasReplacement	isReplacementOf	
	hasEra	-	
hasMethod	isMethodOf		
hasProcess	isProcessOf		

3. 관계 유형의 최적화를 위한 데이터 분석

3.1 분석 대상 데이터

용어 간의 관계가 “용어 X → 용어 Y”의 형식

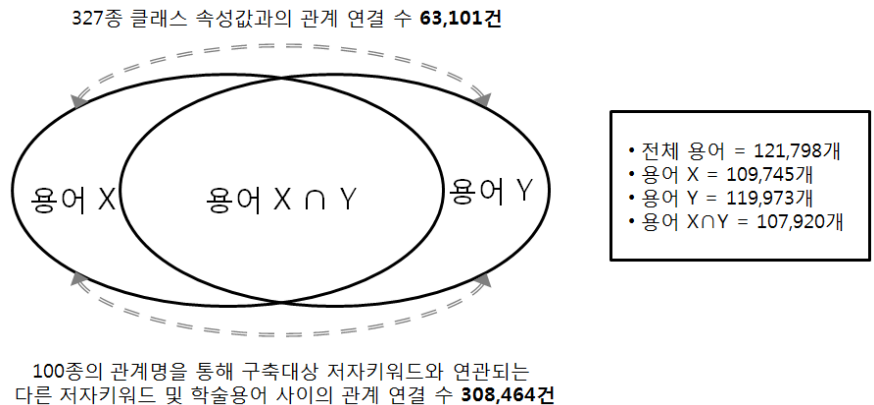
으로 맺어지는 경우, 2014년 12월 16일 현재 STNet에는 용어 X 109,745개와 용어 Y 119,973개 사이에 327종의 클래스 속성에 의한 63,101건의 관계가 형성되어 있으며(〈부록 1〉 참조), 100종의 관계명을 통해 구축대상 저자키워드와 연관되는 다른 저자키워드 및 학술용어들 사이

에 308,464건의 관계가 형성되어 전체 관계 연결 수는 총 371,565건이다(〈그림 2〉 참조). STNet의 입력 대상 저자키워드 6만여 개보다 용어 수가 많은 이유는, 선정된 6만여 개 저자키워드를 해당 클래스의 속성에 따라 정의하는 과정에서 구축자가 관련 용어로서 유의미하다고 판단되는 신규 학술용어를 추가하여 데이터베이스를 구축하였기 때문이다.

STNet의 “용어 X → 용어 Y” 관계를 기준으로 하여 용어군 X에 연결된 관계 수가 많은 순

으로 상위 10개의 클래스를 살펴보면, 「실존인물」에 속하는 용어군 X에 연결된 관계 수가 1위로 41,420개, 이어서 「행위/활동」 12,771개, 「이론/사상」 11,545개, 「정치/법률제도」 11,196개 등으로 나타나고 있다(〈표 3〉 참조).

본 연구의 분석 대상 클래스는 「실존인물」이며, 분석 대상 데이터는 「실존인물」의 용어군 X, X에 연결된 용어군 Y, X와 Y간 연결된 관계이다. 이 중 클래스 「실존인물」에 속하는 용어 X에 연결되는 용어 Y에 클래스가 부여되



〈그림 2〉 STNet의 용어 간 관계 연결 수

〈표 3〉 관계 연결 수 상위 10위에 해당하는 X 용어군의 클래스

순위	클래스 명	클래스 코드	관계 연결 수
1	실존인물	y01-01	41,420
2	행위/활동	b01-01	12,771
3	이론/사상	d01-01	11,545
4	정치/법률제도	d02-02	11,196
5	경제경영제도	d02-03	9,005
6	조직 및 단체명	y06-01	7,620
7	분과학문	d01-03	7,547
8	행위/행동/역할	B01	6,761
9	사회제도	d02-01	6,521
10	문헌명	y02-01	6,452

〈표 4〉 실제 분석 대상 데이터: 용어 X, 용어 Y, X와 Y의 연결 관계

X		X → Y 연결관계	Y	
클래스	용어 수		용어 수	클래스
「실존인물」	1,746개	관계 연결 수: 총 10,371건 - Y(클래스 속성값): 6,587건 - Y(저자키워드 및 학술용어): 3,784건 ※ Y(클래스 속성값): - 속성 8종 (국적, 소속, 시대, 이론/사상, 저작, 직업, 출생지, 출처) ※ Y(저자키워드 및 학술용어): - 관계명 45종	4,197개	118종 ※ STNet 전체 클래스 - 총 177종

지 않은 경우와 검수가 완료되지 않은 경우는 향후 추가적인 보완작업을 거쳐야 하는 불완전 데이터이므로 분석 대상에서 제외하였다. 따라서 본 연구의 실제 분석 대상 데이터는 클래스 「실존인물」에 속한 용어 X 1,746개, 118종의 클래스가 부여된 용어 Y 4,197개, 용어 X → 용어 Y 연결관계 10,371건이다(〈표 4〉 참조).

3.2 연결관계의 유형별 특징

본 연구의 분석 대상인 클래스 「실존인물」의 용어 간 의미적 관계 종류는 크게 둘로 나뉘어진다. 하나는 클래스 「실존인물」에 해당하는 용어 X가 클래스의 속성과 맺는 관계로서 국적, 소속, 시대, 이론/사상, 저작, 직업, 출생지, 출처의 8개 속성과 6,587건의 관계연결이 형성되어 있다. 다른 하나는 클래스 「실존인물」에 해당하는 용어 X가 다른 저자키워드 및 학술용어와 맺고 있는 관계로서 총 45종의 관계명을 통해 3,784건의 관계연결이 이루어지고 있다. 이를 관계의 종류에 따라 관계가 연결된 수의 내림차순으로 정리하면 〈표 5〉와 같다.

〈표 5〉의 관계 연결 수에 있어서 클래스 속성과의 관계 외에 다른 저자키워드나 학술용어와 관계를 맺는 관계 유형 3,784건 중 가장 많이 나타나는 관계명은 RT(1,527건, 40.3%)이다. RT_X, RT_Y 관계명까지 더하면, 그 수는 더욱 많아진다. RT와 RT_X, RT_Y의 구분은 관계의 방향성에 따른 것이다. 용어 X → 용어 Y의 관계가 상호 연관이 있는 경우에는 RT이며, 용어 X를 기준으로 용어 Y에 미치는 연관관계의 경우는 RT_X, 용어 Y를 기준으로 용어 X에 미치는 연관관계의 경우는 RT_Y이다. RT 관계는 용어 간에 관계를 연결해 줄 필요가 있기는 하지만 기존 관계명 중에서는 정확한 관계명이 없다고 판단될 때 적용하는 것이므로 쉽게 사용할 수는 있으나 연관되는 의미가 모호하다. 따라서 RT가 많이 사용되었다는 것은 STNet 시스템 내에서 수백 종의 클래스 속성과 학술용어 간 관계명을 통해 관계가 유형화되고 있음에도 불구하고, 구축과 검수의 전 과정에 사람이 직접 관여하면서 판단하게 되는 관계명의 미세한 차이를 모두 수용하기는 어렵다는 사실을 방증하는 것이다.

〈표 5〉 클래스 「실존인물」의 용어군 X의 의미적 관계 연결

관계 종류	클래스 속성 / 관계명	용어군 Y와의 관계 연결 수	합계
Y: 클래스 속성값 (속성 8종)	직업	1,547	6,587
	시대	1,504	
	저작	1,178	
	국적	899	
	출생지	782	
	이론/사상	431	
	소속	240	
	출처	6	
Y: 저자키워드 및 학술용어 (관계명 45종)	RT	1,527	3,784
	isInstanceOf	541	
	affects	351	
	isAffectedBy	339	
	isMemberOf	243	
	RT_X	134	
	produces	87	
	hasTheory	86	
	RT_Y	82	
	hasIssue	48	
	isFounderOf	48	
	co-occurs-with	42	
	isOppositionOf	35	
	hasOpposition	34	
	isIssueIn	30	
	precedes	20	
	follows	14	
	isSubjectIn	12	
	isOriginOf	11	
	hasProperty	10	
	hasWork	10	
	UF	10	
	hasLocation	9	
	isComponentOf	9	
	hasAgent	7	
	hasFounder	6	
	BT	5	
	manage	5	
	hasPatient	3	
	hasSubject	3	
	originateFrom	3	
	apply	2	
conceptuallyRelatedTo	2		

관계 종류	클래스 속성 / 관계명	용어군 Y와의 관계 연결 수	합계
Y: 저자키워드 및 학술용어 (관계명 45종)	hasAdvocate	2	
	hasEra	2	
	hasPhenomenon	2	
	isConnectedTo	2	
	causes	1	
	hasAntonym	1	
	hasCreator	1	
	hasMember	1	
	hasMethod	1	
	hasProcess	1	
	isA	1	
	isEvaluationOf	1	
	총 합계		

두 가지 관계 종류 중 클래스 「실존인물」의 속성값과 관계를 맺는 데이터의 가장 큰 특징은 관계의 범위와 내용이 명확하고 구체적이라는 점이다. X 용어군은 저자키워드 중에서도 모두 인명(人名)에 해당하는 용어들이며, Y 용어군은 규정된 8종의 클래스 속성에 따라 입력된 객체값이다. 따라서 X 용어군은 '시대'와 '이론/사상'을 제외하고는 대부분 실체(實體)를 갖는 클래스들에 속하는 Y 용어군과 연결관계를 갖고 있다. 예를 들면, '직업'이라는 속성에 입력된 Y 용어군들은 대부분 '직업명'으로 인지될 수 있는 용어들로써 「인간(직업)」, 「인간(지위/벼슬)」, 「인간(역할)」, 「기관/조직」, 「사회단체」 등의 클래스에 속해 있으며, '국적'이라는 속성에 입력된 Y 용어군들 역시 「국가명」 및 「정부(왕조)명」 클래스에 분포하고 있다.

클래스 「실존인물」에 속하는 용어군 X가 연관된 다른 저자키워드나 학술용어와 관계명으로 연결된 3,784건의 관계는 클래스 속성에 의한 관계와는 다른 특징을 갖는다(〈표 5〉 참조).

이 관계는 「실존인물」이라는 클래스가 갖는 특징과는 별개로 구축자들이 용어 X 자체의 개념을 정의하는 과정에서 X가 주요한 학술용어로써 주목 받게 된 이유나 근거에 따라 관련이 있다고 판단한 용어 Y와 맺어주는 관계이다. 이 경우 용어 Y가 속하는 클래스들은 실재적(實在的)인 것이 될 수도 있으나, '사회제도'나 '행위/활동'과 같이 추상적인 클래스에 속할 수도 있다. 또한 「실존인물」의 '저작' 속성이 관계명 hasCreator-hasWork와 의미상 중복될 수 있는 것처럼, 용어군 X의 클래스 속성에서 나타나는 관계 유형과 구축대상 저자키워드 용어 X와 연관된 다른 저자키워드 및 학술용어 Y 간에 형성되는 관계명의 의미가 중복되는 경우도 있다. 이렇게 클래스 속성으로 연결되는 관계 유형과 타 학술용어와 연결되는 관계명의 관계 유형이 의미상 중복되는 경우에는 STNet 데이터 입력 지침에 따라 클래스 속성에 우선해 관계 유형을 연결해주고 있다.

3.3 관계속성 항목 데이터의 특징

저자키워드가 클래스 속성값과 맺어지는 관계에서는 연결의 이유와 내용이 구체적이고 명확하기 때문에 '관계속성' 항목이 사용되지 않는다. '관계속성' 항목은 STNet 구축 입력들이 데이터 입력 과정에서 저자키워드와 다른 저자키워드 또는 다른 학술용어와의 관계명을 부여할 때 정확하게 일치되는 관계명이 없다고 판단되는 경우 사용하는 항목이다. 따라서 '관계속성' 항목의 데이터를 분석할 경우 구축자와 검수자의 의견, 특히 연결해 주는 용어 Y를 선정하고 $X \leftrightarrow Y$ 관계를 정하는 데 있어서의 구축자와 검수자의 논리적인 판단 기준을 알아볼 수 있다.

원칙적으로 STNet 시스템에서는 모든 용어 단위 관계명에서 '관계속성' 값을 입력하는 것이 가능하다. 그렇지만 실제 사용되는 경우를 보면 RT, RT_X, RT_Y에 대부분의 입력값이 몰려 있다. 『실존인물』 클래스의 경우에도 용어군 X와 용어군 Y의 저자키워드 및 타 학술용어와의 연결관계에 적용된 45개 유형의 관계 3,784건 중 절반에 가까운 관계가 RT(1,527건), RT_X(134건), RT_Y(82건)의 세 유형에 몰려 있는 것으로 나타나고 있다(〈표 5〉 참조).

이 중 구축자나 검수자들이 주어진 관계 유형을 좀 더 세분화해 주기 위해 활용한 '관계속성' 항목값은 14개의 관계 유형에 900건의 데이터가 있는 것으로 분석되었다. 14개의 관계 유형에 구축자들이 자유롭게 입력한 '관계속성' 항목값을 중복된 단어를 제외하고 나열하면 〈표 6〉과 같다.

〈표 6〉의 '관계속성' 항목값에서 나타나는 가

장 큰 특징은 인간관계에 대한 내용이 가장 많이 입력되어 있다는 점이다. 그 중에서도 구축자에 따라 부자관계, 부-자, 부/자, 아들, 부자와 같이 조금씩 다른 형태로 표기하고 있으나 가족 또는 친인척 관계에 대한 구분이 압도적으로 많다. 또한 스승과 제자 관계, 창건자, 창간자, 기고자, 출연자, 간행인, 발행인과 같은 사회적인 역할을 나타내는 용어와 실천, 찬성, 추모, 총애와 같은 행위를 표현하는 용어도 많이 나타나고 있다.

두 번째 특징은 동일한 유형에 대한 이형 표기에 있어서 줄표(-)나 빗금(/) 표시 기호를 통해 역관계나 상호작용과 같은 관계의 방향성이 표출되고 있다는 점이다. 예를 들어 아버지와 아들 관계를 표시할 때 본인이 구축하고 있는 용어 $X \leftrightarrow$ 용어 Y의 순서를 염두에 두고 '관계속성' 항목값을 '자-부', 혹은 '자/부'로 표기하거나 '부-자', '부/자' 등으로 구분해서 표기하고 있다.

세 번째 특징은 동일한 유형의 '관계속성' 항목값이 RT, RT_X, RT_Y, affects, isAffectedBy 등에서 나누어 나타나고 있다는 점이다. STNet에서 RT_X와 RT_Y는 방향성을 가지고 있으나 RT는 방향성을 갖지 않으며, affects는 isAffectedBy와 역관계의 의미를 갖는 동시에 RT, RT_X, RT_Y에 비해 '영향력을 주거나 받는' 의미가 구체적으로 담겨 있는 관계명이다. 예를 들어, 스승과 제자 관계가 RT, affects, isAffectedBy의 '관계속성' 항목에 동일하게 입력되어 있다고 하더라도 구축자 입장에서 스승과 제자로써 서로 영향을 주거나 받는 관계가 중요하다고 생각될 경우에는 영향을 주거나 받는 방향에 따라 affects나 isAffectedBy를 입력하고 '관계

〈표 6〉 구축자와 검수자가 입력한 ‘관계속성’ 항목값

연결관계 명	(%)	‘관계속성’ 항목값
RT	(72.89)	계비, 관련 사건, 관련 인물, 관련사건, 관련자, 관련자/사건, 관련제도, 교우, 교우관계, 교유(交遊), 교유관계, 기고자, 기림, 기예능보유자, 남매, 남편, 대표 작곡가, 동료연구자, 동문, 동생, 동성연애대상자, 동인지-동인, 딸, 모/자, 모녀, 모녀관계, 모자, 모자관계, 묘(墓), 문예지/편집자, 문인, 문헌/간행인, 발행인, 배우자, 백부, 번역, 법우(法友), 보유, 보유자, 부, 부/녀, 부/자, 부녀, 부녀관계, 부부, 부자, 부-자, 부자관계, 비, 빈, 사모/대상자, 사위/장인, 사위-장인, 사제, 사-제, 사제관계, 사제지간, 상/수상자, 상사/부하, 생질, 수상, 수상/수상자, 수상명/수상자, 수상자/상, 숙부/조카, 스승, 스승/제자, 아들, 암살, 양자/양부, 연인, 왕, 외삼촌/조카, 외숙/조카, 원작/편저자, 자/부, 자매, 자-모, 자-부, 작품저자, 장서, 장인, 장인/사위, 제-사, 제자, 제자/스승, 제-형, 조부, 조손, 조손관계, 조카, 졸업, 중심 인물, 지양함, 지지, 창간, 창건자-창건대상, 창시, 처음 시도, 첩, 친교, 친구, 친분, 친척, 편집책임자, 편찬, 폐지/주체자, 학문적동료, 할아버지/손자, 헌정, 협판사무(協辦事務), 형제, 후손/조상
RT_X	(14.67)	간행, 개념 생성, 개발, 개작, 계비, 공감(共感), 관련 사건, 관련 양식, 관련 인물, 관련연구, 관심대상, 묘(墓), 반대, 백부, 번역, 보급, 복역, 부도비(浮屠碑), 분류, 비난, 비판, 수상, 수상자/수상, 실천, 암살, 연구대상, 연구및개발, 연출, 오빠, 외교정책, 외증손, 용어 생성, 입교, 자문, 제창, 주창, 주창자, 중도 탈퇴, 중창(重創), 진압, 찬성, 참석자/회의, 참전, 창간, 천거(薦舉), 추대, 추서(追敘), 컬럼 기고, 탐사, 탐험, 발견, 편집, 편집실무담당, 편찬, 평가, 호평(好評)
RT_Y	(9.11)	간행인, 계비, 관련 인물, 기고자, 등장인물, 명예총장, 반대, 배향(配享), 백부, 보유자, 수상, 암살, 외증손, 전사(戰死), 정부(情婦), 제향, 제향(祭享), 주창자, 천거, 천거(薦舉), 첩포, 초대 총장, 총애, 추모, 추배, 추서, 출연자, 충신, 탄핵대상, 편집장, 평가, 합사(合祀)
affects	(0.67)	원형복원, 사제, 설립, 문하생
isAffectedBy	(0.56)	교류 인물, 문하생, 사제, 스승/제자, 조손
UF	(0.44)	묘호, 본명, 세례명, 이칭
co-occurs-with	(0.33)	관련 인물, 2세대
hasIssue	(0.33)	연구자/연구주제, 관련 인물
precedes	(0.22)	제일 1세대/2세대
follows	(0.22)	제일 1세대/2세대
isMemberOf	(0.22)	교사, 본관
hasWork	(0.11)	편저자
isIssueIn	(0.11)	등장 인물
produces	(0.11)	유적/복구한 사람

속성’ 항목값으로 사제관계를 표시한다. 스승과 제자로써의 영향 관계보다는 스승과 제자라는 관계 자체에 중점을 두는 경우에는 RT를 입력하고 사제관계를 표시한다.

마지막 특징은 구축자에 따라 ‘관계속성’ 항

목값으로 입력해 주는 용어의 구체성에 차이가 있다는 점이다. 예를 들어 부모와 자식의 관계를 표현해 주는 경우 부자관계처럼 일반적인 방식으로 기술한 경우가 있으며, 아버지, 아들, 딸, 모자관계, 모녀관계처럼 구체적으로 묘사한

경우가 있다. 또한 관련 인물, 관련 사건, 관련 제도 등과 같이 구체적인 인물 관계나 관계 대상을 묘사하는 데 있어 구체적인 표현으로 구분해주기 어렵거나 못하는 경우에는 ‘관련(關聯)’이라는 매우 모호한 단어를 사용하는 경우도 있다.

4. 관계 유형의 최적화

4.1 최적화를 위한 논리적 기준

‘관계속성’ 항목값의 특징은 크게 네 가지로 분석된다. 첫째, 특정 유형의 용어가 우세하게 많이 입력되어 있으며, 둘째, 관계의 방향성이 제안되고 있고, 셋째, 동일 관계 유형에 대한 이형 표기가 많으며, 넷째, 관계를 표현하는 용어의 의미적 구체성에 차이가 있다. 관계 유형의 최적화는 해당 도메인의 분석 결과에 상응하게 이루어져야 하므로, STNet의 ‘관계속성’ 항목값의 분석에 의한 네 가지 특징에 상응하게 도출될 수 있는 관계 유형 최적화의 기준 또한 네 가지로 정리될 수 있다. 네 가지 기준의 논리와 근거는 다음과 같다.

4.1.1 빈도수에 의한 관계명 신설

STNet은 기존의 관련 연구를 참고해 만든 100개 유형의 관계명을 가지고 있으나, 실제 전문 입력들이 수행한 구축 작업 결과, 전체 입력된 관계 데이터의 20% 이상이 RT 관계에 몰려 있어 관계를 좀 더 구체화시켜 줄 필요가 있다. 특히 「실존인물」 클래스에 주요 학술 용어 중 많은 수가 속해 있음을 고려할 경우, 본 연구

의 분석 결과 ‘관계속성’으로 입력된 값들에 특별히 많이 나타난 유형의 용어에 대해서는 별도의 관계 유형을 범주화해 일부 관계명을 신설하는 것이 바람직하다. 이는 향후 STNet을 활용한 정보서비스를 보다 효율적으로 제공하기 위해서도 필요하다.

4.1.2 빈도수에 의한 관계의 유형화 제한

사실상 같은 의미를 나타내지만 이형으로 표기된 표현들을 모아 범주화하고(예: 부/자, 부-자, 모/자, 모-자 등의 표기는 부모-자식 관계 유형으로 범주화), 범주화한 유형에 실제 ‘관계속성’ 항목값이 많이 입력된 순서로 정리하였다. 그 결과 수백 종의 관계가 나열되어 STNet 구축 작업과 향후 서비스 제공에 있어서 효율성 저하 문제가 있는 것으로 나타났다. 따라서 관계 유형의 출현 빈도수에 따라 빈도가 낮게 나타나는 경우 관계의 유형화를 제한하는 것이 바람직하며, 관계 유형이 지나치게 특정한 용어 관계에만 적용될 가능성이 있는 경우(예: 묘(墓))도 제한할 필요가 있다.

4.1.3 관계의 방향성 고려

출표나 빗금표시 등으로 관계의 방향성이 표현된 ‘관계속성’ 항목값들은 가급적 역관계명을 같이 만들어 향후 시스템에서 활용할 수 있도록 하는 것이 필요하다.

4.1.4 기존에 많이 사용되고 있는 관계명

참조 · 반영

구축자에 따라 동일한 관계 유형을 나타내는 표현이 상이하거나 구체성에 차이가 나는 경우가 많다. 이러한 문제점을 극복하기 위해서는 전 세계

의 주요 RDFS/OWL 기반 프로젝트에서 만든 용어 집들이 탑재되어 있는 Linked Open Vocabularies (LOV, <http://lov.okfn.org>) 등의 용어를 참조하여 구체화 정도의 기준을 정하는 것이 필요하다. 예를 들어, ‘관계속성’ 항목값에서 모자 관계, 부자관계, 모녀관계, 부녀관계 등으로 매우 구체적으로 표현된 유형들은 LOV 관련 용어들의 확인을 통해 부모-자식 관계(isParentOf-isChildOf)를 기준으로 삼아 적용하는 것이 바람직하다.

이와 같은 기준에 따라 본 연구에서는 5회 미만 입력된 표현의 경우 범주화 작업에서 제외하고, 5회 이상 입력된 이형 표기들을 범주화하여 클래스 「실존인물」의 관계 유형을 최적화

하였으며, 그 결과 <표 7>과 같이 세분화된 관계명을 도출하였다.

4.2 세분화한 관계명의 유용성

STNet 「실존인물」 클래스에 연결된 ‘관계속성’ 항목값의 분석 내용을 바탕으로 관계 유형을 범주화해서 신설한 관계명들이 STNet의 RT 관계 구체화에 유용하게 적용될 수 있을지를 확인하기 위해, 실제 STNet의 데이터를 가지고 검증 작업을 수행하였다. 검증을 위한 비교 데이터는 <표 5>의 「실존인물」 클래스에 연결된 데이터 중 RT, RT_X, RT_Y 값을 갖는 1,743건이다. ‘관계속성’ 항목값의 유무에 상관없이 신

<표 7> 관계 유형의 최적화에 따른 관계명의 세분화

관계 유형	관계명	최적화 적용 기준	비고
부모-자식	isParentOf ↔ isChildOf	(1), (3), (4)	혈연으로 맺어진 부모-자식 관계, 법적인 부모-자식 관계(양부모-양자 관계, 결혼에 의한 부모-자식 관계) 포함
배우자	isSpouseOf	(1), (3), (4)	법적 관계 및 사실혼 관계를 모두 포함
조부모-손자	isGrandparentOf ↔ isGrandchildOf	(1), (3), (4)	혈연 및 법적 관계를 모두 포함
형제-자매	isSiblingOf	(1), (3), (4)	혈연 및 법적 관계를 모두 포함
친척	isRelativeOf	(1), (3), (4)	혈연 및 법적 관계를 모두 포함
스승-제자	isMentorOf ↔ isMenteeOf	(1), (3), (4)	-
동료	isColleagueOf	(1), (3), (4)	동료 연구자, 업무 상 동료이자 협력 관계, 친구 관계, 친분 관계를 모두 포함
관련 인물	isRelatedPersonOf	(1), (3), (4)	기타 관련된 인물 관계를 포함하며 특정 문화재, 기관 단체, 사건 등과 관련된 인물인 경우에도 사용할 수 있음
수상(명)-수상자	isAwardedTo ↔ winAward	(1), (3), (4)	-
창작물-편집자	hasEditor	(1), (3), (4)	특정 창작물의 경우에는 저작자(hasCreator) 외에 편집자가 중요한 관계로 연결이 필요할 경우도 있으므로 기존에 있는 hasWork와 연결될 수 있는 hasEditor 관계명을 생성하고, 현재 hasCreator-hasWork 역관계로 정의되어 있는 관계는 단방향 관계로 정리함

* 최적화 기준 '(2)'는 빈도 수가 낮아 관계명을 별도로 신설하지 않는 경우에 적용하였음

〈표 8〉 세분화한 관계명의 유용성

재배정 관계 유형	재배정한 관계명	포괄하는 관계 연결 수	비율
부모-자식 관계	isParentOf ↔ isChildOf	251	14.40%
배우자 관계	isSpouseOf	84	4.82%
조부모-손자 관계	isGrandparentOf ↔ isGrandchildOf	24	1.38%
형제-자매 관계	isSiblingOf	64	3.67%
친척 관계	isRelativeOf	18	1.03%
스승-제자 관계	isMentorOf ↔ isMenteeOf	31	1.78%
동료 관계	isColleagueOf	107	6.14%
관련 인물	isRelatedPersonOf	411	23.58%
수상(명)-수상자	isAwardedTo ↔ winAward	118	6.77%
창작물-편집자 관계	hasEditor	27	1.55%
기존 STNet 관계명으로 구체화 작업이 가능한 경우		145	8.32%
기존에 맺어진 관계 오류로 연결 관계를 삭제한 경우		2	0.11%
신규 관계명 배정을 하지 않은 경우		461	26.45%
총계		1,743	100%

설한 관계명의 재배정 작업을 수행해 기존에 RT, RT_X, RT_Y 관계를 얼마나 포괄할 수 있을 지를 확인하였다. '관계속성' 항목값이 명시되지 않은 RT, RT_X, RT_Y 관계의 재배정 작업은 구축자들이 정리한 용어 개념 정의의 내용을 기준으로 하였으며, 개념 정의에 용어 관계를 구체화 해 줄 내용이 명시되지 않았을 경우에는 신규 관계명을 배정하지 않았다.

재배정 작업 결과, 신설된 관계명이 기존 RT, RT_X, RT_Y의 1,743건 데이터 중 1,135건을 포괄할 수 있는 것으로 나타났다. 이는 기존 관계명으로도 구체화 작업이 가능한 경우나 오류로 인해 관계 연결을 삭제한 경우를 제외한 것으로, 전체 65.12%에 해당하는 관계가 보다 구체화된 관계 유형으로 최적화된 것이다(〈표 8〉 참조). 특히 재배정 작업은 구축 인력이 아니라 연구자가 직접 시스템에 입력된 정의 등의 내용을 검토하여 직관적인 판단이 가능한 수준의 연결 관계에 대해서만 수행하였다. 따라서 구

축 인력을 재배정 작업에 직접 다시 투입해 최적화 작업을 진행할 경우 RT, RT_X, RT_Y 관계를 구체화 할 수 있는 비율이 더 높아질 수 있을 것이다.

5. 결론

본 연구는 KCI에 등재된 인문, 사회, 예술체육 분야 학술논문에서 추출한 학술 용어를 바탕으로 구축하는 구조적학술용어사전 데이터베이스 시스템 'STNet'에 실제 입력된 데이터의 관계들을 분석하여 관계 유형의 최적화와 관계명의 세분화 방안을 찾고자 하였다. 이를 위해 본 연구에서는 STNet의 클래스 「실존인물」에 속한 용어 X 1,746개, 클래스(118종)가 부여된 용어 Y 4,197개, 용어 X → 용어 Y 연결관계 10,371건을 분석하였다. 또한 STNet의 데이터베이스 구축자와 검수자들이 제안하는 관계명 관련 의견

들을 효율적으로 취합하기 위해 만든 ‘관계속성’ 항목의 값을 분석하였으며, 관계 유형 최적화와 관계명의 세분화를 위해 클래스 『실존인물』의 ‘관계속성’ 항목에 입력된 14개 관계 유형의 데이터 900건을 분석하였다.

‘관계속성’ 항목값을 분석한 결과 특정 유형의 용어가 우세하게 많이 입력되어 있으며, 관계의 방향성이 제안되고 있고, 동일 관계 유형에 대한 이형 표기가 많으며, 관계를 표현하는 용어의 의미적 구체성에 차이가 있는 것으로 나타났다. 이러한 특징에 상응하게 관계 유형을 최적화하기 위해서는 빈도수에 의한 관계명 신설, 빈도수에 의한 관계의 유형화 제한, 관계의 방향성 고려, 범형 관계 유형의 반영과 같은 논리적 기준이 필요한 것으로 나타났다. 이와 같은 논리적 기준에 따라 본 연구에서는 5회 이상 입력된 표현을 기준으로 관계 유형을 최적화하고 관계명을 세분화하였다. 그 결과 신설된 10종의 관계 유형에 속한 14개의 관계명을 통해 기존 RT, RT_X, RT_Y의 1,743건의 데이터 중 1,135건의 관계 연결을 구체적으로 표현할 수 있는 것으로 나타났다. 이는 RT, RT_X, RT_Y로 모호하게 연결된 관계의 65.12%가 세분화된 관계명을 가지게 되는 것으로, 실제 시스템에 적용할 경우 입력 및 검색 단계에서 입력자와 검색자에게 상당히 유용할 것으로 해석할 수 있는 결과라 할 수 있다.

사람이 직접 용어 관계를 판단하고 연결해 가는 과정에서 제시되는 다양한 관계 유형의 분석을 통해 신규 관계명을 만들고 세분화하는 최적화 방법은 자동화 기술을 통해 관계를 설정함으로써 사람의 판단을 최소화하고 빠른 시간 내에 효율성을 높이하고자 하는 방법에 비해

구축 속도 면에서는 많이 뒤떨어진다. 그러나 용어 관계의 누적 건수에 비례하여 사람이 직접 판단하는 클래스나 용어 관계들에 대한 양질의 경험적인 데이터가 함께 축적되기 때문에 오히려 자동화 기술을 이용한 프로젝트들이 공통적으로 안고 있는 ‘테스트 데이터’ 부족의 한계성을 극복하는데 도움을 줄 수 있다. 또한 이렇게 구축한 양질의 의미 구조나 관계 유형들을 각종 학술정보 서비스와 연계시킬 경우 학술과 연구 분야의 지식 지도를 형상화할 수 있으며, 연구자들이 중요하게 생각하는 학술적인 의미나 연관 정보들을 추적하고 탐색하는 검색 시스템에 매우 효과적으로 활용할 수 있을 것이다.

본 연구에서 진행한 방식과 동일한 방법론을 STNet 전체 데이터에 적용하여 특정 클래스에 속한 용어군에서 공통적으로 나타나는 관계명을 추출할 경우, 이 관계명은 해당 클래스의 속성 항목으로 추가할 수 있을 것이다. 이런 의미에서 STNet 구축을 위해 선정된 저자키워드가 클래스 속성과 맺는 관계의 경우는 ‘관계속성’ 필드가 없어서 본 연구의 분석에서 제외되었으나, 거시적인 측면에서는 연구 범위에서 완전히 제외된 것은 아니라 할 수 있다.

또한 구체적인 ‘관계속성’ 항목값이 입력되지 않은 RT, RT_X, RT_Y의 경우 신규 관계 유형을 만들기 위한 분석 범위 과정에서는 제외되었으나, 최적화시킨 신규 관계 유형을 재배정하는 검증 과정에서는 ‘관계속성’ 항목값의 입력 유무와 상관없이 RT, RT_X, RT_Y 데이터가 모두 활용되었다. 이와 관련하여 추후 RT, RT_X, RT_Y의 관계 유형으로 연결되어 있는 용어군 X와 용어군 Y가 속한 각 클래스

의 내용을 기준으로 구축된 관계 유형의 사례 연구가 필요할 것으로 판단된다.
들을 유추하고 분석해주는 방식을 이용한 심화

참 고 문 헌

- [1] 고영만. 2006. 시소러스 기반 온톨로지에 관한 연구. 『정보관리』, 5: 5-22.
- [2] 고영만, 김비연, 민혜령. 2014. 한국학술지인용색인(KCI)의 인문학, 사회과학, 예술체육 분야 저자 키워드의 의미적, 형태적 분석에 의한 개념범주 텍사노미 연구. 『한국문헌정보학회지』, 48(4): 297-322.
- [3] 공현장, 황명권, 김원필, 김판구. 2005. 특정 도메인에 대한 자동 온톨로지 구축 방법에 관한 연구. 『한국정보과학회 제32회 추계학술발표회 논문집』, 32(2): 595-597.
- [4] 김은영. 1998. 국어 어휘의 부분-전체관계에 대한 고찰. 『한국어 의미학』, 2: 155-174.
- [5] 김규환, 장보성, 남영준. 2008. 전문용어 기반 인문사회분야 온톨로지 구축에 관한 연구. 『제15회 한국정보관리학회 학술대회 논문집』, 181-188.
- [6] 남영준. 2009. 『재단구축 용어데이터레지스트리의 개념관계 유형 정형화를 통한추론 규칙 적용방안 연구』. 서울: 한국학술진흥재단, 정책연구-2008-007-지식 확산.
- [7] 박옥남. 2011. 정보조직 지식구조에 대한 연구: 2000년~2011년 학술논문을 중심으로. 『한국비블리아학회지』, 22(3): 247-267.
- [8] 백지원. 2005. 『용어관계의 분류 모형 개발에 관한 연구』. 박사학위논문, 이화여자대학교 대학원 문헌정보학과.
- [9] 백지원, 정연경. 2005. 지식조직체계의 용어관계 유형에 관한 연구. 『한국문헌정보학회지』, 39(4): 119-138.
- [10] 성하정, 김장원, 이석훈, 백두권. 2014. 관계형 데이터베이스 구성 요소의 의미 관계를 고려한 RDB to RDF 매핑 시스템. 『정보처리학회논문지: 소프트웨어 및 데이터 공학』, 3(1): 19-30.
- [11] 양창진. 2010. 학술 논문의 주제어 표기 및 활용 방안 연구 - DB 구축 및 정보 연계의 관점에서. 『인문콘텐츠』, 19: 395-416.
- [12] 유영준. 2005. 온톨로지의 개념간 관계 설정을 위한 AGROVOC 시소러스의 분석에 관한 연구. 『정보관리학회지』, 22(1): 125-144.
- [13] 이혜영, 광승진. 2011. 국내 학술지 논문의 주제어를 통한 학술연구분야 관계분석. 『한국비블리아학회지』, 22(3): 353-371.
- [14] 이재호, 김연희, 신현경, 송기봉. 2014. 알츠하이머 관련 논문을 대상으로 하는 온톨로지 기반 지식 표현 방법 연구. 『인터넷정보학회논문지』, 15(3): 125-135.

- [15] 임수연, 박성배, 이상조. 2005. 의미관계 정보를 이용한 약품 온톨로지의 구축과 활용. 『정보과학회 논문지: 소프트웨어 및 응용』, 32(5): 428-437.
- [16] 장령령, 홍현진. 2014. 학술지 중요도와 키워드 순서를 고려한 단어동시출현 분석을 이용한 독서분야의 지적구조 분석. 『한국비블리아학회지』, 25(1): 295-318.
- [17] 전말숙. 1998. 시소러스의 연관관계 유형에 관한 연구. 『정보관리연구』, 29(1): 20-39.
- [18] 정현기, 김유섭. 2008. 도메인 온톨로지 구축을 위한 개념 자동 추출 및 클러스터링. 『한국정보과학회 2008 종합학술대회 논문집』, 35(1): 305-309.
- [19] 정현숙, 최병일. 2005. 텍스트 내용 기반의 철학 온톨로지 구축 및 교육에의 응용. 『정보교육학회논문지』, 9(2): 257-269.
- [20] 조이현, 박대원, 박동훈, 문홍구, 권혁철. 2006. 비전문가에 의한 상하위 관계 중심의 온톨로지 공동구축 방법. 『한국지능정보시스템학회 추계학술대회 논문지』, 2(1): 87-91.
- [21] 조현양, 남영준. 2004. 시소러스와 온톨로지의 상호 호환성에 관한 연구. 『정보관리학회지』, 21(4): 27-47.
- [22] 한성국, 이현실. 2006. 시소러스를 활용한 온톨로지 구축방안 연구: 시소러스의 SKOS 변환을 중심으로. 『한국비블리아학회지』, 17(1): 286-303.
- [23] 황미녕, 이승우, 조민희, 김순영, 최성필, 정한민. 2012. 연구 트렌드 분석을 위한 기술 지식 온톨로지 구축. 『한국콘텐츠학회논문지』, 12(12): 35-45.
- [24] Akbari, Ismail and Fathian, Mohammad. 2010. "A Novel Algorithm for Ontology Matching." *Journal of Information Science*, 36(3): 324-334.
- [25] Maedche, Alexander and Staab, Steffen. 2002. "Measuring Similarity between Ontologies." *of the European Conference on Knowledge Acquisition and Management (EKAW)*, 251-263.
- [26] Zhitomirsky-Geffet, Maayan and Erez, Eden Shalom. 2014. "Maximizing agreement on diverse ontologies with "wisdom of crowds" relation classification." *Online Information Review*, 38(5): 616-633.
- [27] Linked Open Vocabularies. [online] <<http://lov.okfn.org>>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Ko, Young Man. 2006. "A Study on the Ontology based on Thesaurus." *Information Management*, 5: 5-22.
- [2] Ko, Young Man, Kim, Bee-Yeon and Min, Hye-Ryoung. 2014. "A Study on a Conceptual Taxonomy of Author Keywords of Humanities, Social Sciences, and Art and Sport in the

- Korea Citation Index (KCI) by Analysis of its Meaning and Lexical Morpheme.” *Journal of the Korean Society for Library and Information Science*, 48(4): 297-322.
- [3] Kong, Hyunjang, Hwang, Myunggwon, Kim, Wonpil and Kim, Pankoo. 2005. “The Study on the Automatic Ontology Building Methodology about the Specific Domain Knowledge.” *Proceedings of the 32th Fall Conference on Korean Institute of Information Scientists and Engineers*, 32(2): 595-597.
- [4] Kim, En-young. 1998. “A Study on the Part-whole Relation of Korean Vocabulary.” *Korean Semantics*, 2: 155-174.
- [5] Kim, Gyu Hwan, Jang, Bo Seong and Nam, Young Joon. 2008. “A Study on Domain-Specific Ontology Construction by Terminology.” *15th Proceedings of the Korean Society for Information Management Conference*, 181-188.
- [6] Nam, Young Joon. 2009. *A Study on the Inference Rule of Application Conceptual Registry of the Term Data Registry of KRF*. Seoul: National Research Foundation of Korea, Policy Studies-2008-007.
- [7] Park, Ok Nam. 2011. “Knowledge Structures in Knowledge Organization Research: 2000-2011.” *Journal of the Korean Biblia Society for Library and Information Science*, 22(3): 247-267.
- [8] Baek, Ji-Won. 2005. *A Study on the Development of a Classification Model for Terminological Relationships*. Ph.D. diss. Department of Library and Information Science, Ewha Womans University.
- [9] Baek, Ji-Won and Chung, Yeon-Kyoung. 2005. “A Study on the Semantic Relationships in Knowledge Organization Systems.” *Journal of the Korean Society for Library and Information Science*, 39(4): 119-138.
- [10] Sung, Ha Jung, Gim, Jang Won, Lee, Suk Hoon and Baik, Doo Kwon. 2014. “An RDB to RDF Mapping System Considering Semantic Relations of RDB Components.” *KIPS transactions on software and data engineering*, 3(1): 19-30.
- [11] Yang, Chang-Jin. 2010. “Study on Keywords and Their Use of Academic Theses - Focused on Database Development and Information Link.” *Humanities Content*, 19: 395-416.
- [12] Yoo, Yeong-Jun. 2005. “A Study on the Analysis of AGROVOC for Establishment of Concept Relationships of Ontology.” *Journal of the Korean Society for Information Management*, 22(1): 125-144.
- [13] Lee, Hye Young and Kwak, Seung Jin. 2011. “Relation Analysis Among Academic Research Areas Using Subject Terms of Domestic Journal Papers.” *Journal of the Korean Biblia Society*

- for Library and Information Science*, 22(3): 353-371.
- [14] Lee, Jaeho, Kim, Younghee, Shin Hyunkyung and Song, Kibong. 2014. "A Study on Ontology Based Knowledge Representation Method with the Alzheimer Disease Related Articles." *Journal of Internet Computing and Services(JICS)*, 15(3): 125-135.
- [15] Lim, Soo-Yeon, Park, Seong-Bae and Lee, Sang-Jo, 2005. "Medicine Ontology Building based on Semantic Relation and Its Application." *Journal of Korean Institute of Information Scientists and Engineers (KIISE): Software and Applications*, 32(5): 428-437.
- [16] Zhang, Ling Ling and Hong, Hyun Jin. 2014. "Examining the Intellectual Structure of Reading Studies with Co-Word Analysis Based on the Importance of Journals and Sequence of Keywords." *Journal of the Korean Biblia Society for Library and Information Science*, 25(1): 295-318.
- [17] Jun, Mal Suk. 1998. "A Study on the Types of the Associative Relationship in Thesauri." *Journal of Information Management*, 29(1): 20-39.
- [18] Jung, Hyun-Ki and Kim, Yu-Seop. 2008. "Automatic Extraction and Clustering of Concepts for Domain Ontology Construction." *Proceedings of the 2008 Conference on Korean Institute of Information Scientists and Engineers*, 35(1): 305-309.
- [19] Chung, Hyun-Sook and Choi, Byung-Il. 2005. "Building a Philosophy Ontology based on content of Texts and its Application to Learning." *Journal of the Korean Association of Information Education*, 9(2): 257-269.
- [20] Cho, Lee-Hyun, Park, Dae-Won, Park, Dong-Hoon, Moon, Hong-Goo and Kwon, Hyuk-Chul. 2006. "A Study on the Co-Work Methodology for Ontology Hierarchical Relations by non-Experts." *Proceedings of the Fall Conference on Korea Intelligent Information System Society*, 2(1): 87-91.
- [21] Cho, Hyun-Yang and Nam, Young-Joon. 2004. "A Study on the Interchangeability between a Thesaurus and an Ontology." *Journal of the Korean Society for Information Management*, 21(4): 27-47.
- [22] Han, Sung-Kook and Lee, Hyun-Sil. 2006. "A Study of Ontology Construction Using Thesaurus: Transformation of Thesaurus into SKOS." *Journal of the Korean Biblia Society for Library and Information Science*, 17(1): 286-303.
- [23] Hwang, Mi-Nyeong, Lee, Seungwoo, Cho, Minhee, Kim, Soon Young, Choi, Sung-Pil and Jung, Hanmin. 2012. "Ontology Construction of Technological Knowledge for R&D Trend Analysis." *The Journal of the Korea Contents Association*, 12(12): 35-45.

[부록 1] STNet의 클래스 및 객체(object)값을 갖는 클래스별 속성

클래스	클래스 속성	클래스	클래스 속성
[a01-01] 성씨	본관(지명), 시조	[D01] 이론(사상,이념,주의,법칙)	관련자,반대이론,시대,주창자,해당국가
[a01-02-01] 인간(성별)	민족/인종	[d01-01] 이론/사상	관련자,반대이론,시대,주창자,해당국가
[a01-02-02] 인간(연령)	민족/인종	[d01-02] 원칙/법칙	관련자,반대이론,시대,주창자,해당국가
[a01-03-01] 인간(친족관계)	민족/인종	[d01-03] 분과학문	관련자,반대이론,시대,주창자,해당국가
[a01-04-01] 민족/인종/종족	민족/인종	[d01-04] 개념(정의)	관련자,반대이론,시대,주창자,해당국가
[a01-04-02] 국민	민족/인종	[D02] 제도(체제)	대체제도,목적
[a01-04-03] 인간(거주상황)	민족/인종	[d02-01] 사회제도	대체제도,목적,시대
[a01-04-04] 인간(사회계층)	민족/인종	[d02-02] 정치/법률제도	대체제도,목적,시대
[a01-04-05] 인간(세대)	민족/인종	[d02-03] 경제경영제도	대체제도,목적,시대
[a01-04-06] 공동체	민족/인종	[e03-02] 각국어	어족
[a01-05-03] 인간(성향)	민족/인종	[X01] 지명	원지명,위치
[a01-06-01] 인간(직업)	민족/인종, 직업의대상	[x01-01] 대륙(반도)명	위치
[a01-06-02] 인간(지위/벼슬)	민족/인종	[x01-02] 국가명	수도,언어,원지명,위치,인접국
[a01-06-03] 인간(역할)	민족/인종	[x01-03] 도시/구/동명	원지명,위치
[A02] 기관/조직	목적	[x01-04] 산/산맥명	원지명,위치
[a02-01] 행정/공공기관	목적	[x01-05] 바다/강/호수명	원지명,위치
[a02-02] 교육기관	목적	[x02-01] 시대	대상지역,용도
[a02-03] 기업/회사	목적	[X03] 관계/상호작용	대상자,주체자
[a02-04] 사회단체	목적	[x03-02] 비교/차이	대상자,주체자
[a03-01] 동물명	분포지,사례,시대,학명	[x03-03] 계층/등급/계통	대상자,주체자
[a03-02] 식물명	분포지,사례,생지,식물군,학명	[x03-05] 인과	대상자,주체자
[a03-03] 자연물(광물)	사례	[x03-05-01] 원인/조건/요소	대상자,주체자
[A04] 인공물	대체물,보완물	[x03-05-02] 결과	대상자,주체자
[a04-01] 물품/제품/생산품	대체물,보완물	[x03-05-03] 효과/영향	대상자,주체자
[a04-02] 재료/부품	대체물,보완물	[x03-06] 상호작용	대상자,주체자
[a04-03] 교재/자료	대체물,보완물	[x03-06-01] 통합/결합/제휴	대상자,주체자
[a04-04] 의류	대체물,보완물	[x03-06-02] 교환/교류/교제	대상자,주체자
[a04-05] 식료품	대체물,보완물	[x03-06-03] 참여/중재	대상자,주체자
[a04-06] 도구/수단	대체물,보완물	[x03-06-04] 반응/대응	대상자,주체자
[a04-07] 건축물/시설물	대체물,보완물	[x03-06-05] 역(대립/논쟁/투쟁)	대상자,주체자
[a04-08] 교통수단	대체물,보완물	[Y01] 인명	국적,소속,시대,이론/사상,저작,직업,출생지,출처
[B01] 행위/행동/역할	행위대상(자),행위자	[y01-01] 실존인물	국적,소속,시대,이론/사상,저작,직업,출생지,출처
[b01-01] 행위/활동	행위대상(자),행위자	[y01-02] 가상인물	국적,소속,시대,이론/사상,저작,직업,출생지,출처

클래스	클래스 속성	클래스	클래스 속성
[b01-02] 교육활동	행위대상(자), 행위자	[Y02] 창작물명	시대, 원서명, 장르/유형, 저작자
[b01-03] 경제/산업활동	행위대상(자), 행위자	[y02-01] 문헌명	수상명, 시대, 원서명, 이론/사상, 장르/유형, 저작자
[b01-04] 위법행위	행위대상(자), 행위자	[y02-02] 예술작품명	수상명, 시대, 원서명, 이론/사상, 장르/유형, 저작자
[b01-05] 신체운동/행동	행위대상(자), 행위자	[y02-03] 신문/잡지명	시대, 원서명, 이론/사상, 장르/유형, 저작자
[b01-06] 기능/역할	행위대상(자), 행위자	[y02-04] 방송(프로그램)명	시대, 원서명, 장르/유형, 저작자
[B02] 변화(변동)	행위대상자, 행위자	[y02-05] 지도명	시대, 원서명, 장르/유형, 저작자
[b02-01] 완화(감소/축소/쇠퇴)	행위대상자, 행위자	[y02-06] 문서/보고서명	시대, 원서명, 이론/사상, 장르/유형, 저작자
[b02-02] 강화(증가/확장/신장)	행위대상자, 행위자	[Y03] 사건명	결과, 관련자, 대상자, 시대, 원인, 장소, 주관자
[b02-03] 개혁(개편/재편/혁신)	행위대상자, 행위자	[y03-01] 사건명/회담명	결과, 관련자, 대상자, 시대, 원인, 장소, 주관자
[b02-04] 변천/과정	행위대상자, 행위자	[y03-02] 국경일/기념일명	결과, 관련자, 시대, 원인, 장소, 주관자
[b02-05] 교체	행위대상자, 행위자	[y03-03] 행사/축제일명	결과, 관련자, 대상자, 시대, 원인, 장소, 주관자
[b02-06] 분해/합체	행위대상자, 행위자	[y03-04] 수상명	결과, 관련자, 대상자, 시대, 원인, 장소, 주관자
[C01] 특성/성질	대상자	[Y04] 기념물(문화재)명	관리자(처), 소유자(처), 소재지, 제작시대, 제작자
[c01-01] 경향/동향	대상자	[y04-01] 유형물	관리자(처)
[c01-02] 기질/품질/형질/성향	대상자	[Y05] 법률/제도명	시대
[c01-03] 수준/정도	대상자	[y05-01] 법률/법령	시대
[c01-04] 능력/힘/에너지	대상자	[y05-02] 조약/협약	관련국가, 시대
[c01-05] 분포	대상자	[y05-03] 정책(제도)명	시대, 시행국가(지역), 입안자(관련자)
[c01-06] 환경	대상자	[Y06] 기관/단체명	설립자, 소재지, 시대, 이전기관명, 이후기관명
[c01-07] 감각	대상자	[y06-01] 조직및단체명	설립자, 소재지, 시대, 이전기관명, 이후기관명
[C02] 심리	원인, 해당자	[y06-02] 정부(왕조)명	설립자, 소재지, 시대, 이전기관명, 이후기관명
[c02-01] 감정	원인, 해당자	[y06-03] 학파/종파명	설립자, 소재지, 시대, 이전기관명, 이후기관명
[c02-02] 인식/의식	원인, 해당자	[y06-04] 회의명	설립자, 소재지, 시대, 이전기관명
[C03] 현상/이슈	결과, 원인	[Y07] 상품명	사용목적, 시대
[c03-01] 상태/상황	결과, 원인	[y07-01] 도구/수단명	사용목적, 생산국, 시대, 제조자
[c03-02] 격차/차이	결과, 원인	[y07-02] 제품/브랜드명	사용목적, 생산국, 시대, 제조자
[c03-03] 문화/생활	결과, 원인	[y07-03] 건축물/시설물명	사용목적, 생산국, 소재지, 시대, 제조자
[c03-04] 경제/경영/무역	결과, 원인	[Y99] 기타개체명	일반주제어(유개념어)
[c03-05] 정치/국제	결과, 원인		