

구조적 용어사전 구축에서 법률용어 성분분석을 통한 관계속성 지정의 효과에 관한 연구

A Study on the Effects of Application of Relational Attribute on Construction of Structured Glossary - A Semantic Analysis of Legal Terminology

조라현, 성균관대학교 대학원 문헌정보학과, do0825@skku.edu

La-Hyoun, Jo, Library & Information Science, Sungkyunkwan University

본 연구는 법률명 용어의 의미적 성분분석을 통해 ‘보충적 관계속성’을 도출하고, 이를 구조적 용어사전의 관계속성 설정에 적용하였을 때 용어관계에 나타나는 효과를 확인하는 것을 목적으로 한다. ‘보충적 관계속성’이란 클래스(범주)별 용어들이 가지는 특징에 기반하여 만든 관계속성을 말하며, 용어사전 구축 시 도출한 관계속성을 반영함으로써 용어 간 관계가 유기적이고 명확하게 표현될 수 있도록 하는 보조적 구축지침의 역할을 하도록 한다. 이를 위해 법원도서관 ‘법률용어 관련어집’의 316개 용어쌍을 분석하여 ‘보충적 관계속성’을 도출하였으며, 구조적 용어사전 용어에 적용하여 밀도 값과 밀도의 정도를 시각화하여 효과를 확인하였다. 그 결과, ‘보충적 관계속성’ 전, 후의 변화는 적었으나 이는 법률명 용어의 특징으로 인한 결과라는 것을 알 수 있었고, ‘보충적 관계속성’ 반영의 명확한 효과 확인을 위해서는 분석대상용어를 법률명에서 법률관련용어로 확장시켜 연구할 필요성을 확인하였다.

1. 서론

“구조적 용어사전”이란 용어를 구조적으로 정의하는 것으로, 유사한 속성을 가지는 용어를 범주별로 구분하여 개념화하고 개념 범주별 속성 구조에 따라 해당 용어를 정의한다. 이때 용어 간 의미적 연관관계와 속성 간 의미적 연관관계를 설정해 줄 경우 용어의 의미적 추론이 가능한 지식베이스로 발전시킬 수 있다.

구조적 용어사전에서 용어의 의미적 연관관계를 표현하기 위해 용어 간 관계를 맺는 과

정에는 용어의 의미관계를 해(害)하는 몇 가지 문제점이 존재한다. 용어 간에 적절한 관계명이 없어 관계를 설정할 수 없는 경우, 지정된 관계명이 없어 용어사전의 구축자에 따라 상이한 관계 명을 설정하는 경우, 특정 분야의 용어에서 관계명의 의미가 달라지는 경우, 구축자가 지침을 이해하지 못해 다른 관계 명을 설정하는 경우 등이다. 이러한 문제들은 “구조적 용어사전” 내 용어의 의미상 오류로 존재하는 것만이 아닌, “구조적 용어사전”의 의미 확장에도 부정적 영향요소가 될 수 있으며, 추론 가능한 지식베이스로의 발전을 저해

하는 요소로 작용하기 때문에 해결이 필요한 문제라고 할 수 있다. 이에 본 연구에서는 법원도서관의 '법률분야 관련어집'의 용어관계 분석하여 '보충적 관계속성'을 만들고, 이 속성을 구조적 용어사전의 법률명 용어에 적용하여 선(先)구축된 용어를 보완한 후, 용어의 네트워크 분석방법을 통하여 '보충적 관계속성'이 용어의 관계성이 높아지는데 효과가 있는지 확인하는 것을 목적으로 한다.

'보충적 관계속성'이란 클래스(범주)별 용어들이 가지는 특징에 기반하여 만든 관계속성을 말하며, 용어 구축 시 지정된 관계속성을 반영함으로써 용어 간 관계가 유기적이고 명확하게 표현될 수 있도록 하는 보조적 구축지침이라고 할 수 있다. 다만, 본 연구의 결과는 법률명 용어만을 대상으로 하였기에 분석 결과를 법학/법률분야가 아닌 구조적 용어사전의 전주제로 일반화시키기에는 한계가 있다.

2. 선행연구

2.1 용어관계에 관한 연구

문헌정보학에서는 주로 시소러스의 기본관계(BT, NT, RT, USE, UF)로는 용어간 관계를 명확하게 알 수 없는 점을 보완하기 위한 연구들이 이루어져 왔다.

한상길은 1999년과 2000년의 연구에서 기존 시소러스 구축 지침이 사용하고 있는 용어 관계를 분석하고 용어관계들이 완전히 수용되지 못한 점과 관계 설정어의 부족한 부분을 언급하며 범위주기, 정의, 한정어, 이력주기를 유형별로 구분하여 사용함으로써 이용자에게 많은 정보를 제시할 수 있는 부가 용어관계 확장이 필요하다고 하였다. 김지훈, 김태수(2006)는 기존의 시소러스에 표준정의(Df)를 부여하고 구축함으로써 새로이 추가된 관계

중 BT관계의 적용 양상을 분석한 결과 변경된 BT가 47.2%, 새로 부여된 BT가 19.2%로 시소러스 계층수준에 변화를 주었음을 확인하고, 이를 통하여 용어의 의미를 표현하는데 있어 기존의 관계구조만으로는 용어의 본질적인 속성을 이해하기 어려울 뿐만 아니라 그 범위를 한정하기 어려운 점을 언급하며 다양한 주제분야 용어의 표준정의를 작성하기 위한 모델과 정의규칙의 개발이 필요하다고 하였다. 이는 관계명 추가와 타당성 있는 관계 구축이라는 다른 관점에서 용어관계를 보고 있지만 기존의 시소러스의 관계만으로는 용어의 의미를 충분히 반영하기 어렵다는 점을 인식한 연구라고 볼 수 있다.

2.2 시소러스 기반 온톨로지 관련 연구

유영준(2005)는 AGROVOC 시소러스를 분석하여 의미관계를 구조화 하는데 있어 완성도가 부족한 시소러스의 BT, NT, RT, UF 관계를 보완할 수 있는 온톨로지의 개념 간 관계를 제시하였다. 또한 시소러스가 의미 관계 표현에 있어 한계를 가지고 있으나 온톨로지의 개발을 위해서라도 시소러스의 관계에 대한 심층적인 연구가 필요하다고 하였다. 고영만(2006)은 시소러스의 개념 간 관계 구조의 한계 극복을 위하여 온톨로지 형태의 구조를 도입, 변형시키기 위해서는 시소러스 용어들의 관계와 관계의 유형을 분석하여 표준화시켜야 한다고 했다. 유영준과 고영만의 연구는 시소러스의 개념 간 관계의 구조와 의미적 한계 극복을 위하여 온톨로지로의 변환에 대한 연구를 하였으며 이 점은 의미관계에 관한 다른 연구들과 마찬가지로 시소러스의 기본 관계만으로는 용어 간 의미구축에 어려움이 있음을 드러내는 연구라고 볼 수 있다.

황순희, 윤애선(2005)의 연구는 어휘의 의미성분을 고려하여 시소러스 구축이 필요하다

는 점을 제시하며 각각의 용어의 속성을 분석하여 패킷을 구성한 후 용어의 패킷을 바탕으로 속성관계패킷과 관계패킷을 만들어 과학기술분야 시소러스 구축에 관한 연구를 하였다. 정량적인 검증은 없었으나 용어의 의미성분을 고려하지 않았을 때보다 의미성분을 고려하여 시소러스를 만들었을 때 패킷그룹 내에 의미적 거리와 그룹 간의 의미적 거리가 가깝고, 일관성이 존재할 것이라고 언급하였다.

2.3 법률 분야 시소러스 구축 사례

법학 분야에서는 법원도서관의 “법률분야 관련어집”, 국회도서관 표준 시소러스, 한국법제연구원의 법률용어 시소러스 검색시스템 구축 작업이 이루어졌다.

법원도서관의 “법률분야 관련어집”은 법률 분야 문헌의 정보검색 및 색인어 관리에 사용할 목적으로 구축한 표준 시소러스이며, 1998년 처음 발행된 후 2000년, 2007년 개정을 거쳐 2012년에 3차로 개정된 Ver.4까지 발행되었다.

국회도서관의 “표준 시소러스”는 국회도서관에서 사회과학분야 문헌의 정보검색 및 색인어 관리에 사용할 목적으로 구축되었다. 2000년 “국회도서관 시소러스-정치분야”를 시작으로 2001년 경제, 행정, 법률분야 시소러스를 구축하였고, 2002년 정치, 경제, 행정, 법률분야를 제외한 기타 사회과학분야의 시소러스가 구축되었다. 2003년에는 각 분야별 시소러스가 단일체계의 시소러스로 통합되었고 이후 2005년, 2007년에는 기구축된 용어의 관리 및 신규용어 추가의 형태로 관리되고 있다.

한국법제연구원의 “법령정보 시소러스”는 한국법제연구원 법령정보 데이터베이스시스템의 현행법령 정보관리에 사용할 목적으로 구축되었다. 2003년부터 2008년까지 1년 단위

로 구축되었으며 2008년에는 검색프로그램도 함께 구축되었다. 현재는 한국법제연구원의 홈페이지(http://61.81.114.30/the_web/)에서 시소러스 검색서비스를 제공하고 있다. 한국법제연구원의 “법령정보 시소러스”는 주제를 한정하지 않고 현행법령을 구성하는 전 분야를 대상으로 법령정보를 관리하고 검색하기 위해 필요한 용어들을 모아 체계화 하였으며, 법령 검색이라는 목적에 맞게 현행법령에서 사용되지 않는 학술용어, 일상용어 등의 수록은 최소한도로 제한하였다.

3. 연구방법

3.1 용어 선정

성균관대학교 정보관리연구소의 “인문사회 분야 구조적 용어사전 구축 시스템(이하 STNet)”을 위해 선정된 용어 111,797개와 2007년 발행된 “법원도서관 법률분야 관련어집”에 수록된 60,245개의 용어를 정제하여, ‘-법’, ‘-법률’, ‘-령’, ‘-조례’, ‘-규칙’, ‘-조약’, ‘-협정’, ‘-협약’, ‘-의정서’로 끝나는 용어 중 구조적 용어사전에서 구축이 완료된 용어 73개를 분석대상 용어로 선정하였다. 선정된 용어들은 “국민건강보험법”과 같이 법률 공포 번호를 가지고 있는 법률명이거나 법의 분야를 나타내는 “환경법”과 같은 용어이고, 동의관계로 묶인 USE, UF 관계어는 제외하였다(<표 1> 참조).

<표 1> 분석대상용어 추출내용

분 류	전체용어	법률명용어	분석대상용어
STNet	111,797	89	73
법원도서관	60,245	942	

3.2 용어관계 분석

분석대상용어와 관계를 맺고 있는 관계용어 중 use, uf의 동의관계 용어와 외국어를 제외한 관계용어를 분석대상으로 선정하였으며, 이를 제외한 관계용어의 수는 STNet이 699개, 법원도서관이 316개이다(<표 2> 참조).

<표 2> 분석대상 관계용어 추출내용

분 류	분석대상용어	관계용어		
STNet		725	USE, UF	26
			관계어	699
법원도서관	73	594	USE, UF	16
			외국어	262
			BT	57
			NT	50
			RT	209

‘보충적 관계속성’은 STNet과 법원도서관의 분석대상용어와 관계를 맺고 있는 관계용어의 관계속성을 분석하여 구성한다. 관계속성은 법원도서관의 용어관계를 기준으로 하여 법원도서관의 용어집에는 존재하지만 STNet에 용어사전에는 없는 관계속성과 용어를 만드는 것이다. 예를 들어, 분석대상용어 ‘산업재해보상보험법’을 보면 STNet의 관계는 ‘산업재해보상보험법’의 적용대상인 ‘특수고용근로형태종사자’와 상/하위 법률관계인 ‘개별적 근로관계법’과 ‘노동법’이 관계 맺어져 있고, 법원도서관 용어집에서는 상/하위 법률관계인 ‘노동법’, 목적인 ‘산업재해보상’, 대상인 ‘업무상 재해’가 관계 맺어져 있다. 이 중 법원도서관의 ‘산업재해보상’과 ‘업무상 재해’는 STNet에 존재하지 않지만 STNet의 관계명을 활용하여 관계 맺을 수 있는 용어이므로 STNet 용어사전에 용어관계를 맺어주는 것이다(<표 3> 참조).

그러나 ‘보충적 관계속성’을 적용할 때에는

새로운 용어를 생성하지 않고 STNet에 존재하는 용어를 대상으로 하고, 법원도서관의 용어를 STNet의 관계로 변환하여 관계를 맺는 것을 전제로 하였다.

<표 3> 용어관계 분석과정

	분석대상용어	관계명	관계어	용어분석
STNet	산업재해보상보험법	RT	특수고용근로형태종사자	대상
		is Instance Of	개별적근로관계법	예(상/하 법률)
		is Tributary Of	노동법	상/하 법률
법원도서관	산업재해보상보험법	BT	노동법	상/하 법률
		RT	산업재해보상	목적
		RT	업무상재해	대상

이와 같은 방법으로 법원도서관 316개의 관계용어를 분석한 결과 STNet에 보충할 관계는 상/하위 법률, 관련 법률, 법률의 적용대상, 법률 관련 제도, 법률과 관련된 이슈, 법률의 목적, 법률의 특성, 법률 위반 범죄, 해당 법률을 연구하는 학문 분야, 기타 관련어 등으로 도출되었다(<표 4> 참조).

<표 4> 법원도서관 관계용어 속성분석

관계용어 속성	용어수
상위법률	64
하위법률	46
관련법률	36
대상	28
관련제도	27
이슈	16
목적	11
특성	10
범죄	8
학문	1
동의어	1
기타	68

법원도서관 관계용어의 속성을 STNet에 적용하기 위하여 STNet 699개의 관계용어를 분석한 결과 상/하위 법률, 관련 법률, 법률의 적용대상, 법률 관련 제도, 법률의 목적, 법률과 관련된 이슈, 법률의 특성, 법률과 관련된

범죄, 시대, 관련 학문, 기타 관련어로 분류되었고(<표 5> 참조), 이 중 법원도서관의 관계속성 분석 결과와 비교하여 관계가 연결되어 있지 않은 용어 90개를 추가적으로 구축하였다(<표 6> 참조). 다만, STNet의 관계명은 시소러스와 온톨로지의 관계명을 함께 사용하고 있기 때문에 법원도서관 용어집과 일치하는 많은 관계명은 관계의 의미에 따라 적용하였다.

<표 5> STNet 관계용어 속성분석

관계용어 속성	용어수
하위법률	168
상위법률	86
관련법률	74
대상	54
관련제도	47
목적	22
이슈	22
특성	13
범죄	9
시대	5
학문	3
기타	196

<표 6> STNet 보충적 관계속성 적용

관계용어 속성	용어수
대상	25
관련법률	16
상위법률	15
관련제도	15
하위법률	7
목적	4
이슈	4
특성	1
범죄	1
기타	2

4. 분석결과

4.1 밀도 값 변화를 통한 용어관계 변화 분석

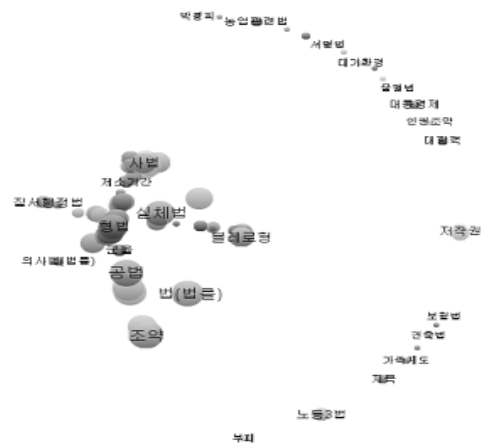
‘보충적 관계속성’ 적용 전, 후의 용어 네트워크 내 연결정도를 알아보기 위해 UCINET을 통해 밀도 값을 측정하였다. ‘보충적 관계속성’을 적용하기 전 밀도 값은 Density (matrix average)=0.0137, Standard deviation=0.1968 이었으며, ‘보충적 관계속성’ 적용 후 밀도 값은 Density (matrix average)=0.0143 Standard deviation=0.1982 이었다.

밀도 값은 0.0137에서 0.0143으로 0.0006만큼 미미한 폭으로 증가하였다.

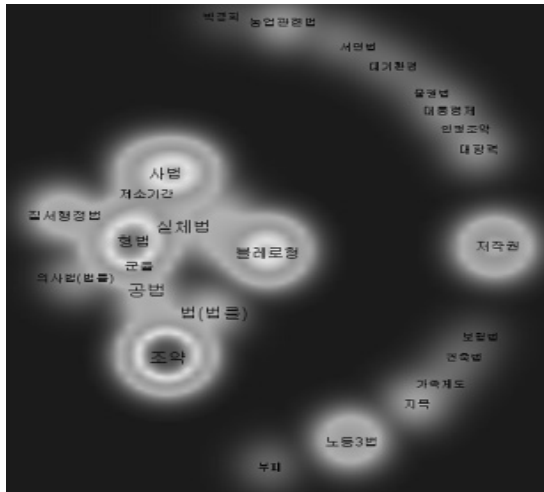
4.2 네트워크 시각화를 통한 용어관계 변화 분석

‘보충적 관계속성’ 적용 전, 후의 용어 밀도의 정도를 VOSviewer를 통해 살펴보았다.

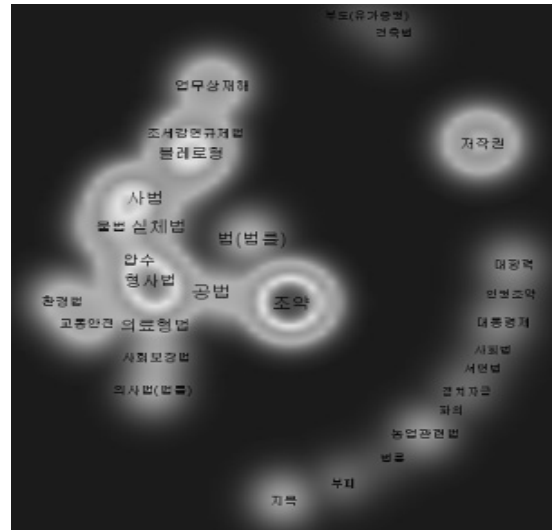
‘보충적 관계속성’ 적용 전의 용어는 총 32개의 그룹으로 군집되었으며, 가장 큰 그룹은 ‘공법’을 중심으로 105개의 용어가 군집되었다. 전체 용어들은 각각의 그룹이 밀집되지 않고 분산되는 형태를 보였다(<그림 1>, <그림 2> 참조).



<그림 1> ‘보충적 관계속성’ 적용 전 STNet의 Label View



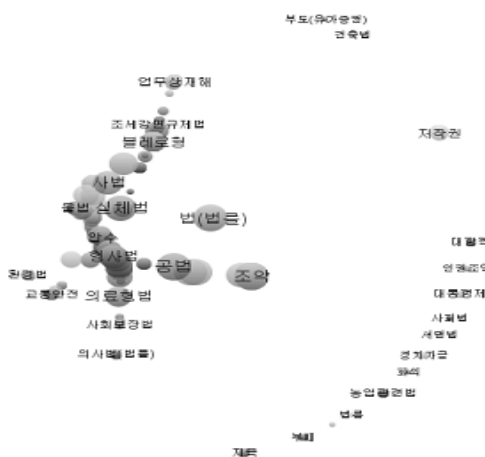
<그림 2> '보충적 관계속성' 적용 전 STNet의 Density View



<그림 4> '보충적 관계속성' 적용 후 STNet의 Density View

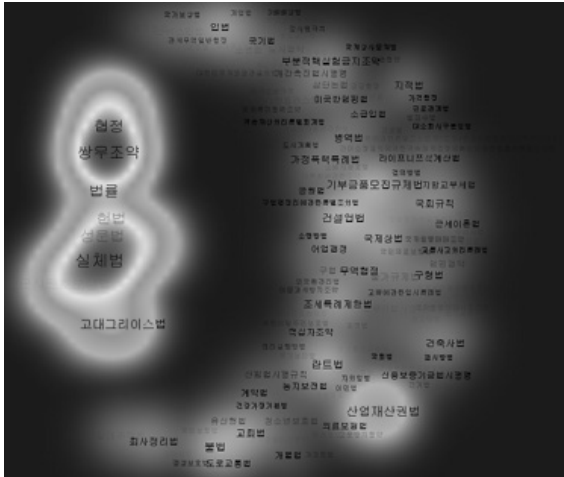
'보충적 관계속성' 적용 후 가장 큰 그룹은 '공법'을 중심으로 99개의 용어가 군집된 그룹이고, 전체 용어는 총 27개의 그룹으로 군집되어 '보충적 관계속성' 적용 전의 32개 그룹 보다 숫자가 줄어 밀집되었다고 할 수 있다. 그러나 군집 형태를 보면 '보충적 관계속성' 적용 전과 마찬가지로 각각의 그룹이 밀집되지 않고 분산된 형태를 보였다(<그림 3>, <그림 4> 참조).

밀도 값 분석과 밀집 정도의 시각화 그림을 통해 변화의 정도가 거의 나타나지 않은 원인이 분석대상 용어로 선정한 표본의 문제인지, 용어의 특성에 기인한 것인지 확인하기 위하여 법원도서관 용어집의 법률명 용어 942개의 관계를 시각화하여 본 결과, 80여개의 그룹이 <그림 2>, <그림 4>와 유사한 형태로 일부만 군집이 되어있는 것을 확인할 수 있었다(<그림 5> 참조).



<그림 3> '보충적 관계속성' 적용 후 STNet의 Label View

위 결과를 법원도서관의 관계용어 속성분석 결과와 관련지어 보면 법률명 용어가 상/하위 법률명과 다수의 관계를 맺고 있기 때문인 것으로 유추해 볼 수 있다. 또한 법률명 용어가 일종의 특수한 전문용어여서 용어사전에서 높은 빈도로 사용될 수 있는 용어가 아니라는 점을 감안할 경우 <그림 1>, <그림 2>, <그림 3>, <그림 4>, <그림 5>와 같은 용어의 밀집형태에 대한 이해도 가능하다. 따라서 이를 바탕으로 분석대상용어를 고립적 관계를 맺고 있는 법률명이 아닌 법률관련 용어로 확장시킨 후 '보충적 관계속성'을 적용하여 그 효과를 확인해 볼 필요성이 있을 것이다.



<그림 5> 법원도서관의 법률명 용어 Density View

5. 결론

본 연구는 STNet 용어사전과 법원도서관 용어집의 법률명 용어를 대상으로 관계어의 관계속성을 분석하여 ‘보충적 관계속성’을 도출하고, 이를 관계속성 설정에 적용하여 용어관계에 나타나는 효과를 확인하였다. STNet과 법원도서관의 관계속성을 분석한 결과 상위 빈도의 관계속성은 하위법률, 상위법률, 관련법률, 대상, 관련제도의 순이었으며, ‘보충적 관계속성’의 적용은 주로 대상, 관련법률, 상위법률, 하위법률을 중심으로 이루어졌다. 밀도 값과 밀집 정도의 시각화를 통해 변화를 확인한 결과 용어 간 밀집에 변화가 있었지만 그 효과는 미미하였다. 법원도서관 용어집의 법률명 용어를 선별하여 용어관계를 시각화한 결과 ‘보충적 관계속성’ 적용 후의 형태와 유사하게 일부만 군집되는 것을 확인할 수 있었다. 이 결과를 법원도서관의 관계용어 속성분석 결과와 함께 살펴보면 ‘보충적 관계속성’의 적용으로 ‘대상’ 관계에 해당하는 일반적 용어들을 법률명 용어에 연결해 주었다 하더라도 관계의 빈도가 높은 상/하위 법률명 용어가 다른 용어들과 다수의 관계를 맺으며 빈도 높게 사용될 수 있는

용어가 아니므로 밀도에 큰 영향을 주지 못한 것으로 분석할 수 있었다.

따라서 본 연구 결과를 바탕으로 분석대상용어를 법률관련 용어로 확장하여 ‘보충적 관계속성’의 적용 효과를 확인하는 후속연구를 진행할 경우 의미 있는 결과를 얻을 수 있을 것이다. 또한 용어관계에 의미 있는 효과가 나타날 경우 ‘보충적 관계속성’을 구조적 용어사전의 법률용어 구축에 반영함으로써 용어 간 관계가 유기적이고 명확하게 표현될 수 있도록 하는 보조적 도구로서 활용할 수 있을 것이다.

참고문헌

- 고영만. 2006. 시소러스 기반 온톨로지에 관한 연구. 『정보관리』, 5: 5-22.
- 고영만. 2012. 2012년도 토대연구지원사업 연구 계획서.
- 국회도서관. 2007. 『(정보검색을 위한) 표준 시소러스』. 제1권-제5권. 서울: 국회도서관.
- 김지훈, 김태수. 2006. 용어정의와 관계추출을 통한 시소러스 확장에 관한 연구. 『한국문헌정보학회지』, 40(1): 293-314.
- 법원도서관. 2007. 『법률분야 관련어집』. 서울: 법원도서관.
- 유영준. 2005. 온톨로지의 개념간 관계 설정을 위한 AGROVOC 시소러스의 분석에 관한 연구. 『정보관리학회지』, 22(1): 125-144.
- 한국법제연구원. 『법령정보 시소러스』. [online]. [cited 2014. 1. 24]. <http://61.81.114.30/thesa_content.asp>.
- 한상길. 1999. 시소러스의 기본 용어관계에 관한 연구. 『한국도서관·정보학회지』, 30(4): 107-136.
- 한상길. 2000. 시소러스의 부가관계에 관한 연구. 『정보관리학회지』, 17(2): 119-138.
- 황순희, 윤애선. 2005. 의미적 준거의 세목화를 고려한 과학기술 분야 시소러스 구축. 『한국어어미학』, 18: 99-124.
- UCINET ver.6. <<https://sites.google.com/site/ucinetsoftware/downloads>>.
- VOSviewer ver.1.5.7. <<http://www.vosviewer.com/download/>>.