

구조적 학술용어사전 데이터베이스 구축에 있어서 용어의 의미관계 형성에 영향을 미치는 요인에 관한 연구*

A Study on the Factors Influencing Semantic Relation in Building a Structured Glossary

권 선 영 (Sun-Young Kwon)**

목 차

- | | |
|-------------------|------------------------|
| 1. 서론 | 5. 가설 검증을 위한 실험 데이터 분석 |
| 2. 이론적 배경 및 선행 연구 | 6. 가설 검증 |
| 3. 연구 절차 및 방법 | 7. 결론 |
| 4. 실험 시스템 및 데이터 | |

초 록

본 연구는 구조적 정의에 의한 학술 용어 사전 데이터베이스 구축을 위해 학술용어의 의미관계 형성에 미치는 요인이 무엇인지를 찾아내고 이러한 요인이 어떠한 영향을 미치고 있는지를 밝히고자 하였다. 이를 위해 2007년부터 2011년 사이에 한국연구재단 등재학술지에 등재된 학술논문에서 학술논문 주제어를 추출하여 이를 대상으로 주제복합성, 언어 네트워크 특성, 출현빈도, 출현패턴을 분석하고 구축된 STNet의 의미관계 형성정도인 용어의 의미적 연결관계 노드의 수와 유형의 수와의 영향 관계를 살펴보았다. 가설 검증을 통해 구조적 학술용어사전의 구축에 있어 의미관계 형성정도에 주요한 영향을 미치는 요인으로 매개 중심성, 출현빈도, 구조적 공백성의 효과크기라는 것을 알 수 있었다. 그리고 용어의 중요성은 일반적으로 알려진 출현빈도를 통한 방법 외에도 연결정도 중심성, 근접 중심성, 매개 중심성, 위세 중심성과 같은 측정방법에 따라서는 판단할 수 있음을 확인하였다. 또한 주제복합성은 직접적으로 의미관계 형성정도에 영향을 미치지 않지만 용어의 근접 중심성에 영향을 미치기 때문에 크게 4가지의 요인을 고려하여 용어를 선정할 경우 의미관계 형성정도는 높아질 수 있는 것으로 파악 되었다. 본 연구의 결과는 지금까지 용어를 선정하는 프로세스의 주된 방법론인 용어의 출현빈도를 활용하는 방법 이외에도 용어 네트워크상에서의 용어의 위치나 주제복합성 같은 방법론을 적용하여 용어를 선정할 수 있다는 것을 보여준다. 따라서 전문용어 사전을 구축할 때 용어의 네트워크에서의 매개 중심성, 출현빈도, 구조적 공백성의 효과크기, 용어의 주제복합성을 면밀히 판단하여 다각도로 용어를 선정할 경우 전문용어 사전의 질적인 향상과 완성도가 높아질 것을 기대할 수 있다.

ABSTRACT

The purpose of this study is to find factors to affect on the formation of semantic relation from terminology and what is to be affected by these factors to build the database scheme of terminology dictionary by a structural definition. In this research, 826,905 keywords of 88,874 social science articles and 985,580 keywords of 125,046 humanities science articles in the KCI journals from 2007 to 2011 were collected. From collected data, subject complexity, structural hole, term frequency, occurrence pattern and an effect between the number of nodes and the number of patterns which were derived from the semantic relation of linked terms of established 'STNet' System were analyzed. The summarized results from analyzed data and network patterns are as follows. Betweenness Centrality, term frequency, and effective size affect the numbers of semantic relation node. Among these factors, betweenness centrality was the most effective and effective size. But term frequency was the least effective. Betweenness Centrality, term frequency, and effective size affect the numbers of semantic relation type. Term frequency is the most effective. Therefore, when building a terminology dictionary, factors of betweenness centrality, term frequency, effective size, and complexity of subject are needed to select term. As a result, these factors can be expected to improve the quality of terminology dictionary.

키워드: 용어사전, 학술용어, 네트워크분석, 지적구조

Glossary, Terminology, Network Analysis, Intellectual Structure

* 이 논문은 성균관대학교 대학원 박사학위 논문을 축약·정리한 것임.

** 성균관대학교 정보관리연구소 선임연구원(freesson@gmail.com)

논문접수일자: 2014년 4월 16일 최종심사일자: 2014년 5월 7일 게재확정일자: 2014년 5월 22일

한국문헌정보학회지, 48(2): 353-378, 2014. [http://dx.doi.org/10.4275/KSLIS.2014.48.2.353]

1. 서론

1.1 연구의 필요성

학술논문 주제어는 제목이나 초록과 더불어 해당 논문의 성격과 내용을 파악할 수 있게 한다는 점에서 학술논문의 핵심 요소가 된다. 학술논문 주제어는 일상적으로 사용되는 일반 용어와는 달리 전문용어(terminology)로 이루어져 있다. 전문용어는 특정 분야에서 새로운 개념, 사물을 구체적으로 지시할 필요에 의해 생긴 것이기 때문에 형태적으로 보다 세분화되어 있고 내용적으로도 구체적인 의미를 표현하고 있다는 특징을 지니고 있다. 또한 전문용어는 어휘 특정성(lexical specificity)이 크며(조은경 2001), 많은 언어적 정보를 담고 있기 때문에 특정 분야 내에서 개념화하는 것이 용이하다.

이러한 전문용어들은 대체로 연구자나 분야 전문가에 의해 학문 분야의 커뮤니케이션 수단인 학술논문에서 주로 사용되며 구체적으로 학술논문 주제어를 통해 나타난다. 학술논문 주제어는 용어가 세분화되고, 구체적 의미를 갖는 전문용어의 특징을 지니고 있기 때문에 개념의 기본적 내용과 성질이 동일한 학술 논문 주제어는 특정 개념범주로 개념화하여 그 개념 범주가 갖는 공통된 속성들을 통해 정의하는 것이 가능하다. 고영만 외(2013)는 이렇게 정의된 사전을 '구조적 학술용어사전'이라 하였으며, 개념범주가 갖는 속성들을 통해 연관관계를 맺어줄 경우 해당 전문 용어가 수록된 연구 성과들의 의미적 연결망을 구성하는 것이 가능해진다고 하였다.

시소러스에 비해 용어 간의 관계성이 더욱

강조되는 구조적 학술용어사전을 구축하기 위해서는 수집된 학술용어를 범주별로 분류하고, 또 범주별로 구조화된 속성에 따라 의미를 정의해야 한다. 이때 용어가 가지고 있는 용어자체의 의미적 특성에 대한 분석 및 용어와 다른 용어들 간의 의미관계 형성에 대한 분석이 이루어질 경우 구조적 학술용어사전에서의 용어의 의미관계를 보다 용이하게 설정할 수 있다. 하지만 의미관계의 용이한 설정을 위해서는 적합한 용어가 선정되어야 한다.

학술논문 주제어는 대체로 해당학술논문의 주제를 반영하고 있는 특징을 지니고 있으며, 특정 용어들은 다른 학문분야에서도 자주 사용되는 '주제복합성'을 띠고 있다. 용어의 주제복합성 정도는 용어의 분야 특정성을 나타내는 것으로 학문 간의 관계성과 용어간의 관계성에 연관되어 있다. 따라서 용어간의 의미관계 정도는 용어의 전체 네트워크에서 해당 용어가 네트워크의 중심에 위치하는 정도 즉 네트워크 중심성을 통해 살펴볼 수 있으며, 네트워크 구조에서의 용어의 위치가 보여주는 구조 내지는 공백의 분석을 통해서도 파악할 수 있다. 그리고 학술용어의 관계를 연결하는 네트워크 내에서의 용어의 관계 수에 따라서도 용어의 중요성을 파악할 수 있다.

학술논문의 주제어(저자키워드)는 또한 그 출현빈도가 높을수록 의미적 연관성이 풍부할 것이며(김관준, 이재윤 2012; 김희정 2011; 서선경 2013; 양창진 2010), 해당 논문의 피인용 횟수와도 상관관계를 가진다(고영만 외 2013). 학술논문 주제어는 1회 출현하는 용어에서부터 수백 회까지 출현하는 용어까지 그 누적출현빈도가 다양하다. 또한 어떠한 학술논문 주제어

는 시간의 흐름에 따라 생성되거나 소멸되는 일종의 생명주기를 갖고 있어서 시기별로 균등하게 출현하기도 하고 특정 시기에만 출현하기도 한다. 따라서 학술논문 주제어의 누적 출현빈도와 시기별 출현빈도 분석을 통한 출현패턴을 살펴보면 용어 간 의미관계의 정도와 범위 파악이 가능하다. 그리고 구조적 정의에 의한 학술용어사전은 용어의 생명주기를 반영함으로써 용어의 최신성 및 용어 간의 유기적인 관계를 유지할 수 있을 것이다.

용어의 주제복합성과 출현빈도 등이 용어 간의 의미관계형성에 미치는 영향을 파악하게 될 경우 용어 간의 관계성 파악을 통해 사전 구축 대상 용어를 효율적으로 선정할 수 있게 된다. 그러나 용어 간의 의미관계 형성과 관련하여 용어의 주제복합성과 출현빈도 등이 의미관계 형성에 어떠한 영향을 미치는지에 대해서는 지금까지 구체적으로 연구된 바 없다. 그리고 구조적 학술용어사전 구축을 위한 용어의 선정 방식 또한 구축자와 도메인에 따라 다르며 표준적 방법이 제시되어 있지 않다. 따라서 용어 간의 의미관계를 형성하는 데 영향을 미치는 요인을 밝히는 연구는 구조적 정의에 의한 학술용어사전 구축에 매우 중요하다.

1.2 연구의 목적

구조적 정의에 의한 학술용어사전 데이터베이스 구축에 있어서 용어 간 의미형성에 영향을 미치는 요인을 찾아내어 용어 선정에 적용할 경우 보다 다양한 의미적 연관 관계를 가지는 용어의 선정이 가능하다. 그리고 구축된 데이터베이스 또한 풍부한 지식의 의미망을 갖게

되어 그 효율성과 활용성이 높아질 것이다. 본 연구의 목적은 실험 대상이 되는 구조적 용어사전의 용어 관계와 관계 유형 및 용어의 주제복합성, 의미관계 네트워크, 출현빈도, 출현시기를 학문 분야별로 분석하고, 구조적 정의에 의한 학술 용어 사전 데이터베이스 구축에 있어서 학술용어의 주제복합성, 의미관계 네트워크에서의 위치, 출현빈도 및 용어출현패턴이 의미관계 형성에 어떠한 영향을 미치고 있는지를 밝히는 데 있다.

2. 이론적 배경 및 선행 연구

2.1 선행 연구

본 연구와 관련한 선행 연구들은 용어의 형태적 특성과 용어의 네트워크 분석과 관련한 연구를 근간으로 한 연구들과 이를 기반으로 한 용어사전 및 구조적 사전의 구축과 관련한 용어의 비통제어회적 특성, 시소러스와 관련한 연구들이다. 용어의 의미, 의미관계를 분석하고 확장하기 위한 다양한 선행연구의 특징은 다음과 같다.

학술논문 주제어들은 전문용어로 이루어져 있으며 낮은 모호성과 높은 특정성 때문에 지식 도메인을 개념화하거나 용어 간의 관계 형성이 용이한 특징을 지니고 있다. 학술논문 주제어는 의미적 특수성을 표현하기 위하여 다양한 합성어의 형태로 존재하는데 이러한 구조를 통해서 의미 관계를 살피는 연구가 발표된 바 있으며(조은경 2001; 주홍군 2010; 이민우 2011; 조개나 2012) 이러한 조어적 형태, 특성을 바탕으로

로 이를 활용한 연구 또한 살펴볼 수 있다. 이수진(2012)은 '국어국문학' 분야의 핵심어를 추출하여 특성을 파악하고 분야에 대한 어휘 선정의 추출에 대한 객관적 분석 방법을 제시한 바 있다. 이처럼 용어의 특성 분석은 사전을 구축하는데 있어 중요한 요소라는 것을 살펴볼 수 있다(박정오 2000; 강정미 1999).

다음으로 용어의 네트워크 분석은 용어와 용어간의 관계를 분석해 볼 수 있는 분석방법으로써 연구주제의 동향, 연구 분야 및 현상에 대한 지식과 이해를 살펴보고자 시도되기도 하였다. 이는 문헌정보학, 행정학, 정치학, 신문방송학, 간호학 등 다양한 분야에서 활용되고 있는 것을 살펴볼 수 있다(김일환 2011; 이수경 외 2011; 조재인 2011; 박옥남 2011; 최영철 2011; 정성훈 2012). 이러한 연구들은 과거의 연구들이 단순한 관계와 출현빈도들에 중심을 두었던 것에 비해 용어는 실제 사용된 맥락을 중심으로 관계가 분석된다는 것에 관심을 두었던 것이 차이점이라고 할 수 있다. 대체로 언어 네트워크 분석을 위한 분석기법 중 동시출현단어 분석방법이 많은 연구에서 사용되고 있는 것을 살펴볼 수 있는데 키워드의 동시출현빈도를 계산 분석하여 어떤 문제를 해결 할 수 있는 잠재력을 지니고 있기 때문이다.

학술논문 주제어는 저자에 의해 부여되기 때문에 비통제어휘의 특징을 지니고 있다. 이에 대해 연구자들은 학술논문 주제어의 형태적, 내용적 특성에 대해 분석을 한 후 이를 활용하기 위해서는 표준화가 필요하다는 언급을 한 바 있다(양창진 2010; 김관준 외 2012).

용어에 대한 정의를 통해 표현된 사전과 용어 간의 관계를 정의한 시소러스는 그 구조적

인 한계를 지적받아 온 것이 사실이다(남영준 1994; 강성희 외 1995). 이에 대한 제약을 해결하기 위해 용어의 분석을 통한 관계의 확장 및 자동화된 구축, 그리고 최근에는 온톨로지로의 확장 등의 연구를 살펴볼 수 있다(강성희 외 1995; 김태수 2001; 최석두 2002; 김완중 2003; 조현양 2004; 유영준 2005; 고영만 2007; 송유화 2012).

마지막으로 용어 사전이나 시소러스 또한 생성시점의 용어를 반영하기 때문에 최신용어를 대응하기 힘든 문제가 제기 되어 왔던 것을 살펴 볼 수 있다. 이러한 문제를 해결하기 위해 연구자들은 어떠한 용어라도 생성과 소멸에 있어 주기를 가지고 있음을 살펴보고 이를 통해 용어 사전이나 시소러스의 최신성에 대한 보완을 고려하였다(최석두 2006; 김은영 2010; 김일환 외 2011).

이상의 선행 연구들을 살펴본바 용어의 의미와 의미관계를 확장하기 위해 기본적으로 용어의 형태분석이 이루어진 것을 알 수 있고 이를 통해 시소러스와 같은 구조적 사전의 한계를 벗어나려는 시도가 있었음을 살펴볼 수 있다. 용어의 네트워크 분석은 용어들 간의 관계를 통해 대체로 연구의 동향을 살펴보는데 이용되었을 뿐 이를 통해 용어의 의미를 확장하는 시도는 이루어진 적이 없는 것을 볼 수 있다. 용어의 생명주기에 대한 연구도 용어가 주기를 갖고 있다는 것을 밝혀내기는 하였지만 이에 따른 활용이나 의미의 부여와 같은 문제는 다루어지지 않은 것을 볼 수 있다.

이와 같은 방법들은 용어의 의미와 용어 간의 의미관계를 살펴볼 수 있는 유용한 방법이지만 각각의 방법 또는 한 가지 이상의 방법을

동시에 적용한 연구들은 국내외적으로 살펴볼 수 없으며 만약 이러한 방법들을 적용할 경우 어떤 효과가 있을 지에 대한 가능성을 살펴볼 필요성이 있다.

3. 연구 절차 및 방법

3.1 연구의 설계 및 절차

본 연구에서는 구조적 정의에 의한 학술용어사전의 데이터베이스 구축에 있어서 의미관계 형성에 영향을 미치는 주요 요인들에 관한 이론, 선행연구 및 실험 대상 데이터에 관한 분석을 바탕으로, 의미관계 형성요인이 의미관계 형성에 어떠한 영향을 미치는지를 밝히고자 한다. 연구의 목적을 달성하기 위해서 다음과 같은 절차와 방법으로 연구를 진행하였다.

첫째, 용어의 의미관계 형성 요인 도출: 용어의 의미관계에 영향을 미치는 요인을 도출하기 위하여 연구의 이론적 배경인 전문용어의 형태 및 특성, 용어의 네트워크 특성과 선행 연구인 동시출현단어에 관한 연구, 저자 키워드에 관한 연구, 용어 사전 관련 연구, 용어생명주기에 관한 선행 연구 분석을 통해 용어의 의미관계 형성에 미치는 요인을 도출하였다.

둘째, 가설 설정: 용어의 의미관계 형성에 영향을 미치는 주요한 요인으로 도출된 주제복합성 정도, 용어의 의미관계 네트워크 구조에서의 위치, 학술지에서의 용어의 출현빈도에 대한 가설을 설정하였다.

셋째, 실험 데이터 구축 및 보정: 성균관대학교 정보관리연구소에서 2012년도 한국연구

재단의 토대연구 지원을 받아 수행 중인 『인문사회 분야 연구성과물의 지식지도 형성을 위한 구조적 용어사전 지식베이스 구축』연구에서 구축한 “STNet(Structural Terminology Net)”을 본 연구의 실험 시스템으로 사용하였다(URL: <http://www.stnet.re.kr> 또는 <http://www.structureterm.org>). 실험을 위한 데이터는 STNet 구축을 위해 수집된 학술논문 주제어와 구축된 구조적 학술용어사전의 용어이다. 다음으로 학술논문 주제어의 주제복합성분석, 네트워크 분석, 빈도분석을 수행하기 위하여 데이터 범위의 제한, 용어의 형태적, 의미적 표준화 작업 등의 데이터 보정 작업을 실시하여 실험 데이터로 사용하였다.

넷째, 데이터 분석: 데이터 보정 작업이 적용된 학술논문 주제어를 분석하여 주제복합성 측면, 네트워크 측면, 빈도 측면에서 어떠한 특성을 보이고 있는지 살펴보았다. 또한 구축된 학술용어사전의 분석을 통해 용어의 의미적 관계가 어떻게 형성되었는지를 살펴보았다.

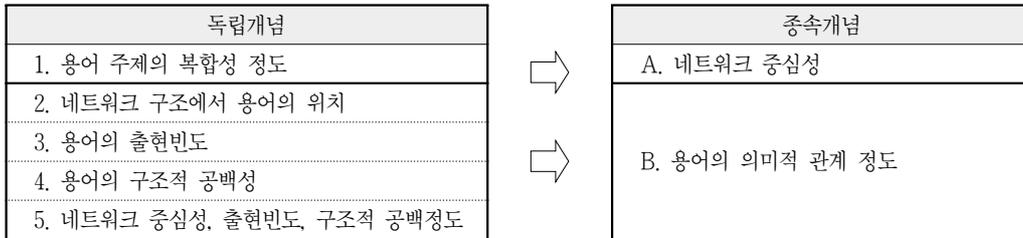
다섯째, 가설 검증: 연구의 가설 검증을 위해 상관관계 분석, 회귀 분석, 일원변량분석통계 방법을 이용하였다.

3.2 연구 가설

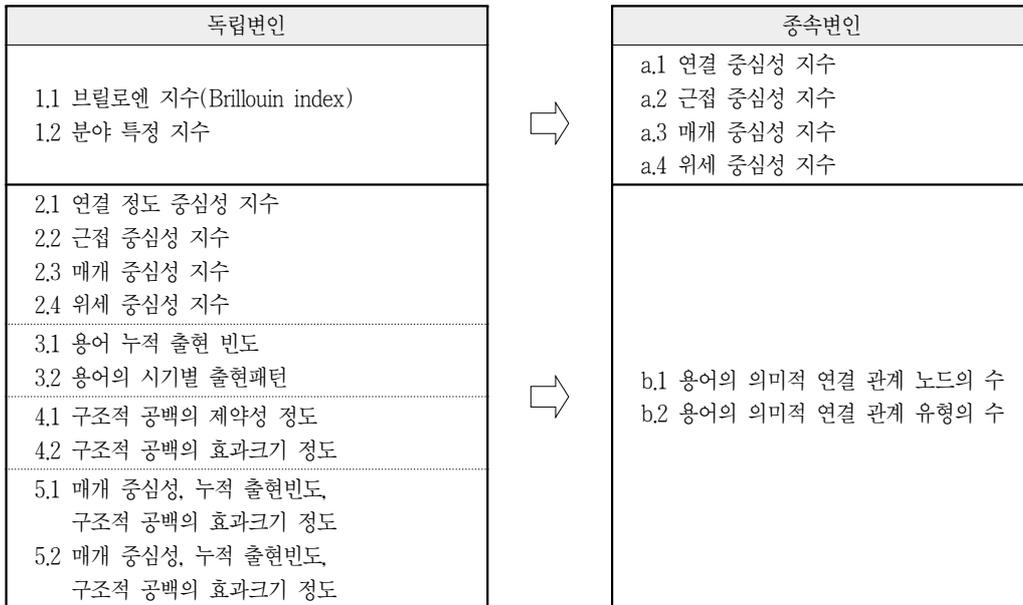
본 연구의 목적은 학술용어의 주제복합성, 네트워크의 위치, 출현빈도와 용어의 생명주기가 반영된 용어의 시기별 출현패턴이 용어 간의 의미관계 형성에 미치는 영향을 밝히는 것이며, 본 연구의 목적을 이루기 위한 가설은 아래와 같다. 가설에 따른 변인과의 관계는 <그림 1>, <그림 2>와 같다.

- 가설 1: 용어의 의미관계 네트워크 구조에서 용어의 주제복합성은 네트워크의 중심성 정도에 영향을 미칠 것이다.
- 가설 2: 용어의 의미관계 네트워크 구조에서 용어의 위치는 용어의 의미관계 형성 정도에 영향을 미칠 것이다.
- 가설 3: 학술지에서의 용어의 출현빈도는 용어의 의미관계 형성 정도에 영향을 미칠 것이다.

- 가설 4: 용어의 의미관계 네트워크 구조에서 용어의 구조적 공백성은 용어의 의미관계 형성 정도에 영향을 미칠 것이다.
- 가설 5: 학술논문 주제어의 네트워크 중심성, 출현빈도, 구조적 공백성은 용어의 의미관계 형성 정도에 영향을 미칠 것이다.



<그림 1> 개념적 가설의 도식화



<그림 2> 조작적 가설의 도식화

3.3 변인 및 측정

3.3.1 용어의 주제복합성 변인 측정 도구

학술논문 주제어는 해당학술논문의 주제를 반영하고 있는 특징을 지니고 있다. 하지만, 용어의 특성에 따라 해당학문의 특성을 반영하는 경우도 있고, 특정용어의 경우 해당 학문 이외에도 다른 학문분야에서도 나타나는 것을 볼 수 있다. 이를 본 연구에서는 '용어주제의 복합성'이라 정의하고 이를 측정하기 위해 브릴로엔 지수(Brillouin index)¹⁾와 분야특정용어지수(Subject-oriented term Index: SOTI)를 활용하였다. 분야 출현 비율은 용어의 전체 출현빈도에서 해당 분야별 빈도의 비율로써 브릴로엔 지수는 지수의 수치가 높을수록 학제성이 높은 것으로 파악되지만, 분야 특정 용어 지수의 경우 수치가 높을수록 '분야특정성'을 가지는 특성을 지닌다. 본 연구에서는 이렇게 '분야특정성'을 지니는 용어를 분야특정용어(Subject-oriented term: SOT)로 정의 하며 이를 산출하기 위해 분야특정용어 지수(Subject-oriented term Index: SOTI)를 제안하였다(〈그림 3〉 참조).

3.3.2 네트워크 중심성 변인 및 측정 도구

사회네트워크 분석에서는 분석하고자 하는

대상을 중심성 분석을 통해 전체 네트워크에서 중심에 위치하는 정도를 살펴볼 수 있다. 중심성 분석은 계량서지학 분야에서 인용 분석을 통해 도출한 네트워크 속에 가장 중심적인 저자, 문헌, 용어 등을 분석하는 데 주로 사용되어 온 바 있다(김현수 2010). 본 연구에서는 분석하고자 하는 대상을 학술논문 주제어로 정의하고 연결 정도 중심성 지수, 근접 중심성 지수(closeness centrality), 매개 중심성 지수(between centrality), 위세 중심성 지수(Eigenvector centrality)를 Ucinet으로 측정하였다.

3.3.3 용어의 출현빈도 변인 및 측정 도구

2007년부터 2011년에 출판된 학술논문의 주제어를 대상으로 해당 년도마다 출현한 용어의 횟수를 합산하여 산출하였다. 다음으로 용어 시기별 출현 패턴을 산출하기 위해 해당 시기의 출현 횟수를 산정한 후 분석을 통해 패턴을 산출 하였다.

3.3.4 용어의 구조적 공백 변인 및 측정 도구

용어의 구조적 공백 변인은 '제약성(Constraints)'과 '효과크기(Effsize: effective size of ego's network)'이다. 먼저 '제약성'은 미국의 사회학자인 Burt(1992)에 의해 주장된 개념인

산출 공식	구분		
	$\frac{n_i}{N}$	분야 특정 용어 지수	N
	n_i		해당 분야에 속한 용어의 수

〈그림 3〉 용어의 주제복합성 분석을 위한 분야 특정 용어 지수 산출 공식

1) 이 지수는 Shannon 외(1949)가 제시한 정보이론의 개념인 '다양성'을 기반으로 하고 있다. 다양성이 높으면 불확실성(Uncertainty)이 높고 다양성이 낮으면 불확실성이 낮은 기본 개념을 바탕으로 한다.

데 구조적 공백 입장에서 가장 중요한 효과는 정보 확보의 우월성을 살펴볼 수 있다(Burt 1992). 일반적으로 사회 네트워크에서는 네트워크의 연결을 행위자로 보지만, 본 연구에서는 이를 용어로 보고 구조적 공백 정도 제약성 정도를 Ucinet을 통해 산출 하였다 다음으로 '효과크기'는 각 용어가 갖고 있는 관계의 수에서 연결된 용어가 갖고 있는 관계의 수의 평균을 뺀 값이다. 즉 하나의 용어가 관계를 맺고 있는 용어가 적은 수의 관계수를 유지할 경우 그 용어의 네트워크 효과크기가 커진다는 것을 의미한다. 구조적 공백의 제약성과 마찬가지로 구조적 공백의 효과크기는 Ucinet을 통해 산출 하였다

3.3.5 용어의 의미적 연결 관계 형성 정도 변인 및 측정 도구

용어의 의미적 연결 관계 형성 정도 변인은 '관계 노드의 수'와 '관계 유형의 수'이다. 용어 관계 노드의 수에서 노드는 용어가 맺고 있는 다른 용어이며 노드의 수는 용어가 맺고 있는 용어의 의미적 연결 관계의 수를 의미한다. 이를 측정하기 위해서 구축된 구조적 학술용어사전에서 각 용어의 관계 용어수를 산출 하였다. 용어의 의미적 연결 관계 유형의 수는 관계가 성립된 용어의 유형 개수를 의미한다. 이를 측정하기 위해서는 실제 구축된 용어 내에서 관계 용어를 추출한 후 그 유형을 구분한 뒤 개별 개수를 산출하였다.

4. 실험 시스템 및 데이터

4.1 STNet의 의미관계 개념 설계

구조적 학술용어사전구축은 연구자들이 사용하는 다양한 표현의 주제어를 수집하여 이들의 관계를 정의하고 다른 개념들과의 관계를 정립하는 일이 필수적이다. 이를 위해서는 기본적으로 관련 개념들을 군집시킬 수 있는 개념의 범주화가 필요하며 중하위 개념이 계층 구조를 이루어 전개될 수 있어야 한다. 그리고 분류된 용어는 정의와 속성을 부여하되, 표현의 다양성, 다의성, 계층 관계의 다중성, 개념의 부분 공유, 중첩 구조, 관계의 상대성 등에 관한 내용을 담아내는 것이 중요하다. STNet의 구조는 크게 용어기술 부분과 클래스 설정 및 개념속성 기술 부분, 관계속성의 기술 부분으로 구성되어 있다. '학술용어의 개념추출', '개념의 속성 도출 및 상세화', '기본관계 및 관계속성의 정의' 과정을 거쳐 의미 관계 개념을 설계하였다.

4.2 STNet의 의미관계 데이터베이스 구축

구조적 정의에 의한 학술용어사전의 데이터베이스 구축에 있어 가장 먼저 선행되는 것은 데이터 수집이다. 이를 위해 STNet 구축 작업에서는 한국연구재단 한국학술지인용색인(KCI) 등재 학술논문 수록 저자키워드를 수집하였다. 다음으로 데이터 수집시의 오류, 특수문자, 괄호, 외국어, 오/탈자, 공백, 불용어 등을 교정하는 형태적 표준화 과정과, 단/복수, 동의어, 조사 등의 의미적 표준화 과정을 통해 용어 정제를 진행하였다. 용어의 표준화 과정 후에는 이

$$set(x) = [[set(F_a) \geq TR] \cap [set(SF_b) \geq SR]] - set(C)$$

$set(x)$ = 선정 용어 집합
 $set(F_a)$ = 전체키워드내의 용어출현횟수(a)이상의 집합
 TR = 전체키워드 빈도 Rank
 $set(F_b)$ = 분야키워드내의 용어출현횟수(b)이상의 집합
 SR = 분야키워드 빈도 Rank
 $set(C)$ = 중복용어, 일반용어, 불용어 등의 군

〈그림 4〉 용어 선정을 위한 산출식

를 바탕으로 구조적 사전 구축을 위한 용어를 선정하는 과정을 진행 하였다. 먼저 앞서 형태적, 의미적 표준화 과정을 거친 후 1회 출현하는 용어는 선정과정에서 제외하였다. 이는 용어의 중요성 측면에서 큰 의미를 둘 수 없기 때문이다(고영만 외 2013). 다음으로 출현빈도를 통해 용어를 선정하고자 하였는데, 일부 특정 분야가 많이 선정 되는 현상을 발견하였으므로 이를 해결하고자 약 5만 용어에서 2회 이상 출현하는 용어 중 학문 분야 별 비중을 계산하여 분야별로 가중치를 두어 선정하였다(〈그림 4〉 참조).

데이터베이스 구축은 구축을 위한 프로그램과 보조 도구를 통해 이루어졌다. 일관성 있는 구축을 위해서 용어 선정, 용어 기술 원칙, 클래스 및 속성 설정 원칙, 관계 및 속성 기술 원칙, 시스템 입력 규칙을 포괄하는 구축 지침서를 작성하여 배포하였다. 또한 입력 교육을 포함한 정기적 구축 회의를 통해 설계자, 구축자, 시스템 개발자가 오류 및 개선 사항을 논의하였으며, 논의 결과를 통해 나온 수정 사항과 보완 사항을 지속적으로 시스템과 구축 작업에 반영하였다.

4.3 실험을 위한 데이터 보정

본 연구의 연구 모형에 대한 가설 검증을 위해서는 수집된 학술논문 주제어에 대한 분석과 구조적 용어사전에 대한 분석이 필요하다. 하지만 형태적, 의미적으로 비 표준화된 특징을 그대로 지니고 있으므로 이러한 문제들을 해결하기 위해 첫째로 형태적, 의미적 표준화 과정을 진행하여 용어를 정제하였다. 두 번째로 분석 주제 분야를 제한하고 다음으로 용어의 규모(개수)를 제한하였다. 세 번째로 분석 방법에 맞도록 데이터를 보정하였다.

5. 가설 검증을 위한 실험 데이터 분석

본 연구의 연구 모형에 대한 가설 검증에 앞서 본 연구의 실험 대상이 되는 학술 논문 주제어가 기초 통계 및 네트워크 측면, 빈도 측면, 용어의 주제복합성 측면에서 어떠한 특성을 보이는지를 분석하였다. 또한 구축된 구조적 학술용어사전의 분석을 통해 용어의 의미적 관계

가 어떻게 형성되었는지를 분석하였다. 실험 데이터에 대한 네트워크 측면, 빈도 측면, 용어의 주제복합성 측면의 분석 결과 값을 통해 가설의 검증을 하였다.

5.1 학술 논문 주제어 분석

5.1.1 기초 통계 분석

본 연구는 한국연구재단 등재학술지에 등재된 2007년부터 2011년도 사이의 인문학 분야 학술논문 88,874건의 학술논문 주제어 826,905건과 사회과학 학술논문 125,046건 중 985,580건의 학술논문 주제어를 기초 분석 대상으로 하였다. 기초 통계를 위해 먼저 인문학, 사회과학 분야에 해당하는 논문 수를 살펴보고 해당 분야 하위의 주제로 논문 수의 분포와 해당 키워드수를 살펴보았다. 구조적 사전에서는 구축 시 인문학과 사회과학 각 4개의 분야와 예술체육의 2개의 분야 총 10개의 핵심 분야를 선정하였다. 하지만 예술, 체육 분야는 본 연구에서 제외하였으므로 인문학의 역사, 철학, 언어학, 한국어와 문학 4개 분야 그리고 사회과학의 법학, 경제학, 경영학, 교육학 4개 분야 총 8개 분야를 분석대상으로 하였다. 분석 데이터가 되는 학술논문 주제어 또한 이 8개의 분야로 제한하여 분석을 수행하였다.

학술논문 주제어에 대한 일반적인 현황을 인문학과 사회과학으로 나누어 분석해 본 결과 논문 수나 키워드 수 모두 두 분야에서 특정 학문분야로 편중되어 있는 것을 볼 수 있다.²⁾ 개

별 논문수의 키워드 수를 살펴보면 특정학문 분야의 경우 다른 분야에 비해 더 많은 키워드가 부여되어 있는 것을 살펴볼 수 있으며 이는 학문의 분야별 특성을 반영하고 있는 것을 의미한다.

5.1.2 주제복합성 분석

학술 용어의 주제복합성을 분석하기 위하여 브릴로엔 지수와 전체분야 대비 해당 분야 출현 빈도의 데이터를 분석하였다. 먼저 브릴로엔 지수는 59,121건의 용어를 대상으로 산출하였다. 결과를 살펴보면 평균 0.30과 유사한 값(0.3056)인 용어는 '북한법', '실질환율', '근대국어'와 같은 용어였는데 이러한 용어들은 2개, 또는 3개 정도의 분야에서만 나타나는 용어이며 주제 분야의 특정성을 지니고 있는 용어라고 할 수 있다. 평균값인 0.30 보다 아래로 내려갈 경우 용어의 출현분야는 더 적어진다. 즉 최솟값인 0이란 수치는 용어가 출현하는 분야가 단 1개임을 의미한다. 반면 지수 값이 높아질수록 출현하는 분야는 더 다양해지는 것을 살펴볼 수 있다. 지수 값의 상위 25%의 용어 수는 3,494개로써 평균 1.9289의 지수 값을 가지고 있다. 반면 지수 값의 하위 25%의 용어 수는 46,235개로써 0.2996의 지수 값을 가진다. 이처럼 지수 값의 하위 25%의 용어 개수가 상위 25% 개수에 비해 13.2배가 넘는 것은 지수값 0의 개수가 37,180개로 전체 59,121개의 62%를 차지하고 있기 때문이다. 이러한 분포는 학술 용어의 62%는 주제의 다양성을 지니기 보다는 분야 특정성을 지

2) 인문학 전체의 키워드의 분포를 살펴보면 이중 상위 20%에 해당하는 5개의 분야에서 전체 64%의 키워드수를 차지하고 있는 것을 살펴볼 수 있다. 사회과학의 경우 상위 20%에 해당하는 5개의 분야에서 전체 56%의 키워드수를 차지하고 있는 것을 살펴볼 수 있었다.

니고 있다고 볼 수 있다.

두 번째로 분야특정용어 지수(SOTI)를 분석하였다. 브릴로엔 지수는 값이 낮을수록 주제복합성이 없고 특정주제에서만 출현하는 성향을 지니는 것에 비해 분야특정용어 지수는 값이 높을수록 주제복합성이 없고 특정주제에서만 출현하는 성향을 지닌다. 분야특정지수를 살펴본 결과 각 학문 분야마다 분야를 반영하는 용어의 비율의 차이가 있음을 알 수 있다. 이는 지수 값이 1인 경우이며 이와 같은 경우 용어가 분야 외에서는 전혀 출현하지 않은 것을 뜻한다(〈표 1〉 참조).

5.1.3 네트워크 분석

본 연구에서 네트워크 분석의 핵심은 중심성 분석, 구조적 공백의 값을 분석하여 이러한 변인이 의미관계 형성정도에 미치는 영향을 살펴보는 것이다. 네트워크 분석은 인문학의 역사, 철학, 언어학, 한국어와 문학 4개 분야 그리고 사회과학의 법학, 경제학, 경영학, 교육학 4개 분야 총 8개 분야로 나누어 수행하였으며 각 분야에서의 중심성 분석과 구조적공백의 정도를 산출하도록 하였다. 연결정도 중심성 값이 낮은 용어를 살펴보면 대부분 저 빈도의 용어가 나

타나는 것을 볼 수 있는 반면 중심성 값이 높은 용어 중에는 고빈도의 용어만 나타나는 것이 아니라 저 빈도의 용어가 나타나기도 한다. 예를 들어 ‘역사학’ 분야에서 ‘표류민’과 같은 용어는 출현빈도가 7회 정도인데, 중심성 값은 4.441로 상당히 높은 편이다. 연결정도 중심성 지수는 국지적인 지역 중심성을 측정하는데 이와 같은 용어는 이를 중심으로 하는 용어 내에서 상당히 중심성이 높다고 해석해볼 수 있다(〈표 2〉, 〈표 3〉 참조).

근접 중심성의 경우 일반적으로 근접 중심성의 값이 높은 경우 전체 네트워크에서 중심을 차지하고 있는 용어라고 할 수 있다. ‘역사학’ 분야의 경우 이에 해당하는 용어는 ‘중국’(15.901), ‘조선’(15.897), ‘고구려’(15.789) 등이다. 반면 근접 중심성의 값이 낮은 용어는 ‘파문’(11.156), ‘부역노동’(11.18), ‘경제협력’(11.406)으로써 이러한 용어는 네트워크의 중심에서 상당히 멀리 떨어져 있는 용어라고 할 수 있다. 매개 중심성은 네트워크 내에서 다른 용어들 사이에 위치하는 정도를 의미하는데 최솟값이 0인 경우가 적지 않다. 이와 같은 용어들은 직접적으로 두 개의 용어 간에서만 관계가 이루어졌다는 것을 살펴볼 수 있다.

〈표 1〉 분야특정지수에 따른 용어 구분

분야	분야	지수	개수	분야내 비율
인문학	역사학	1	596	16.7%
	철학	1	811	22.1%
	언어학	1	146	11.2%
	한국어와 문학	1	1,982	23.3%
사회과학	법학	1	6,926	50.4%
	경제학	1	531	18.4%
	경영학	1	9,864	55.5%
	교육학	1	3,213	34.0%

〈표 2〉 역사학 분야 출현빈도 5회 이상의 용어 리스트

순위	용어	빈도	순위	용어	빈도
1	고구려	145	21	동학	41
2	백제	128	22	근대성	39
3	신라	90	23	냉전	37
4	민족주의	77	24	기억	36
5	한국전쟁	70	25	고조선	36
6	조선	69	26	조선총독부	36
7	중국	67	27	민주주의	34
8	임진왜란	66	28	이주	34
9	발해	59	29	독도	34
10	역사교육	57	30	왜구	34
11	역사교과서	53	31	러일전쟁	34
12	대한민국임시정부	52	32	불교	33
13	정체성	51		.	
14	고려	51		.	
15	일본	51		.	
16	고종	50	1419	코뮌	5
17	3.1운동	49	1420	퇴계학과	5
18	이승만	46	1421	파리의방전교회	5
19	근대화	43	1422	한국노총	5
20	동아시아	42	1423	토지	5

〈표 3〉 연결정도 중심성 지수 리스트

순위	용어	연결정도 중심성	순위	용어	연결정도 중심성
1	고구려	8.770	21	역사교과서	4.006
2	백제	7.480	22	임진왜란	3.971
3	신라	6.585	23	근대화	3.971
4	조선	6.366	24	송환	3.965
5	민족주의	5.885	25	한성	3.958
6	일본	5.192	26	마한	3.938
7	국사	4.802	27	러일전쟁	3.928
8	중국	4.726	28	역사교육	3.919
9	대한민국임시정부	4.721	29	고려	3.914
10	사회주의	4.676	30	유교	3.883
11	동아시아	4.58	31	동북공정	3.878
12	국민	4.417	32	조공체제	3.874
13	표류민	4.411		.	
14	고조선	4.344		.	
15	동아일보	4.240		.	
16	정체성	4.193	1419	펠로폰네소스전쟁	0.085
17	동학	4.190	1420	해월	0.095
18	민족	4.124	1421	토지조사사업	0.123
19	고분	4.090	1422	조흔	0.126
20	청일전쟁	4.044	1423	서체	0.135

위세 중심성은 중심성이 높은 용어와 연결 될 경우 위세 중심성이 높아진다. '역사학' 분야에서 위세 중심성이 가장 높은 용어는 '표류민'(37.485)이었는데 이는 '송환'(30.266) '조선'(23.134)과 같이 위세 중심성이 높은 용어와 연결이 되어 있기 때문이다.³⁾

구조적 공백의 제약성을 살펴보면 최댓값은 1 최솟값은 0.019 평균 0.21이다. 구조적 공백의 제약성 값이 1이라는 것은 용어가 다른 한 개의 용어와만 관계가 맺어져 있는 것을 의미한다. 예를 들어 '파문'이란 용어는 '교회법'이란 용어와만 관계를 맺으며 제약성 값은 1이다. 반대로 제약성 값이 낮다는 것은 용어 네트워크에서 유리한 위치에 있음을 의미한다. 구조적 공백의 제약성 값이 가장 낮은 용어는 '민족주의'(0.019)로 즉 이는 역사학 네트워크에서 용어의 연결을 위한 매우 중요한 용어라는 것을 알 수 있다. 구조적 공백의 효과크기는 용어가 맺고 있는 용어가 다른 용어와 적은 수의 관계 수를 유지할 경우 그 용어의 네트워크 효과크기가 커지는 것을 의미하며 '고구려', '백제', '신라'와 같은 용어가 이에 해당된다.

이와 같이 네트워크 분석을 수행하면서 발견된 점 중 하나는 여러 분야에서 동시에 출현하는 용어에 대한 네트워크상에서의 위치 비교의 필요성이다. 예를 들어 한국어와 문학 분야의 일부 용어는 언어학 분야에서도 동일하게 출현하는 것을 볼 수 있었기 때문에 그 네트워크의 차이가 어떠한지 대략적으로 확인을 해보고자 하였다. 다만 각 네트워크에서 비슷한 위치의 용어를 비교하지 않을 경우 그 편차가 상당히

클 것을 감안하여 각 분야에서 상위에 있는 용어 중 관계용어가 동일하게 출현하는 비율이 높은 용어를 선정하였는데 이는 '한국어교육'이다. '한국어교육'과 같은 용어는 언어학 분야에서 출현빈도 37회로 2위에 해당되며 한국어와 문학 분야에서는 출현빈도 378회로 1위에 해당되는 용어이다. 두 분야는 네트워크의 크기가 다르기 때문에 절대적으로 비교하는 것은 무리가 있지만, 해당 용어를 중심으로 각 분야의 네트워크의 구조를 대략적으로 살펴볼 수는 있다. '한국어교육'의 용어를 중심으로 한 각 분야 네트워크를 비교해 보면 언어학 분야는 관계 용어수가 69개로 103개인 한국어와 문학에 비해 네트워크의 규모가 작으며 관계 수 또한 27% 적은 규모이다. 이는 상대적인 값이므로 네트워크를 비교할 수 있는 밀도를 살펴보아야 하는데 이 값 또한 언어학은 0.070으로 한국어와 문학 분야의 0.117에 비해 작은 편이지만, 분야 전체 네트워크로 비교했을 때의 밀도 차에 비해서는 큰 차이는 아니라고 할 수 있다(언어학 분야 네트워크의 밀도는 0.006, 한국어와 문학 분야의 0.02로 상당히 큰 밀도차가 있다). VOSviewer를 통해 클러스터링을 해보면 언어학은 34개의 클러스터, 한국어와 문학 분야는 12개의 클러스터로 구분되는데 그만큼 언어학 분야의 용어들은 응집되어 있기 보다는 흩어져 있는 네트워크로 분석된다. 네트워크에서 관계 용어중 상위 출현빈도의 용어를 살펴보면 언어학은 '활용형', '명사', '연어', '텍스트구조', '화용정보', '대조언어학'과 같이 언어학적 관점에서 더 가까운 용어들이 나타나고 한국어와 문학에

3) 위세 중심성 값으로 정렬하여 순위를 매겨보면 1,423개의 용어 중 '송환'은 4번째 '조선'은 10번째에 해당하는 용어이다.

서는 '교육내용', '학교문법', '외래어', '통합교육', '스토리텔링' 과 한국어교육이나 문학 분야에 더 가까운 용어들이 나타난다. 이처럼 하나의 용어를 통해 다른 분야에서의 네트워크를 비교해보면 용어를 중심으로 하는 군집 결과나 관계성을 통해 네트워크의 지적구조를 대략적으로 살펴볼 수 있다. 용어가 분야의 대표성을 지닌다는 가정을 해보면 이는 상당히 의미 있는 분석이 될 수 있다(〈표 4〉 참조).

5.1.4 출현빈도 분석

출현빈도에 대한 분석은 구축을 위해 선정된 분야와 분야에 속한 키워드를 대상으로 진행하였는데, 논문수와 키워드의 수를 기준으로 인문학과 사회과학에서 다수 비중을 차지하는 분야를 선정한 후 용어를 선정하였다. 용어의 평균 출현빈도는 8.89이며 총 출현빈도는 488,744이다. 기술 통계량을 살펴보면 분산이 597.2로 상당히 크며 왜도(Skewness) 또한 13.27로 원

쪽으로 상당히 치우쳐진 값을 알 수 있다. 이와 같은 현상의 원인 중 하나는 54,969개의 용어 중 출현빈도가 낮은 용어들이 상당히 많기 때문이다. 출현빈도 2회의 용어만 해도 34%의 비중을 차지하며 출현빈도 3회를 포함하면 이는 이미 50%를 차지한다. 즉 수의 용어가 다수의 출현빈도를 가지고 반대로 다수의 용어는 소수의 출현빈도를 가지는 현상을 보여준다.

5.1.5 출현시기별 분석

2007년 인문학 분야의 키워드출현빈도는 최솟값 0, 최댓값 47, 중앙값 1, 평균 2.32, 표준편차 3.47로 나타났다. 왜도와 첨도는 4.67과 34.23으로 나타나 정규분포를 형성하지는 않는다. 2008년부터 2011년까지도 수치의 차이는 있으나 분포상 큰 차이를 보여주고 있지는 않다. 또한 사회과학 분야의 키워드 출현빈도를 살펴봐도 인문학 분야와의 분포상의 큰 차이점은 발견되지 않는다. 주의 깊게 살펴보아야 할 점

〈표 4〉 용어에 대한 분야 네트워크간 비교

용어	분야	관계용어	관계용어(공통)	용어수	관계수	밀도
한국어 교육	언어학	활용형 명사 기능문법 언어 텍스트구조	대화표지 의미기능 코퍼스 텍스트 문학교육 어휘교육 문법교육 문화교육 교재 한국어교재 보편성 교육과정 다문화가정 한국어학습자 학문목적한국어 텍스트유형	69	338	0.070
	한국어와 문학	매체언어 교육내용 불안 다의어 학교문법 국어교육	외래어 통합교육 읽기교육 표준어 스토리텔링	103	1,236	0.117

은 전체출현 누적결과에서도 살펴보았듯이 여전히 일부 용어의 출현빈도가 상대적으로 높고, 이로 인해 분포상 첨도가 높아 정규분포를 따르지 않고 있다는 점이다. 다만 용어의 전체출현 누적 결과를 분석했을 때는 멱함수 분포를 따르고 있었음에 반해 용어의 시기별 출현 횟수는 그러한 분포를 따르지 않는 차이점이 있다. 하위 학문주제별로 살펴본바 평균적인 키워드 증가율에 비해 상대적으로 상당히 높거나 낮은 분야들이 있는 것을 알 수 있다. 이와 같은 분야들은 통계량만으로는 그 원인을 알 수가 없으며 이러한 원인을 살펴보기 위해서는 개별 용어의 시기별 출현빈도를 살펴보고 출현빈도를 통해 용어의 시기별 출현패턴을 도출해 볼 수 있다.⁴⁾ 이러한 용어의 시기별 출현빈도는 최소자승법을 통해 기울기 값과 결정계수를 구할 수 있다.⁵⁾ 이와 같은 방법을 통해 각 학문분야의 기울기 값을 살펴보면 인문학분야의 기울기 값의 평균값은 0.36이며 결정계수 값은 0.31이다. 사회과학분야는 기울기 값의 평균은 0.40이며 결정계수 값은 0.32로써 큰 차이는 없는

것으로 나타났다.⁶⁾ 하지만 언어학과 같은 분야는 기울기 값이 0.51로 평균 0.36보다 높지만 결정계수 값은 0.29로 평균 0.31과 큰 차이가 없다. 반면 법학과 같은 분야는 기울기 값은 0.29로 평균 0.32보다 약간 낮지만 결정계수 값은 0.29로 평균 0.39에 비해 비교적 낮은 값을 알 수 있다. 이와 같은 분석을 통해 보면 학문분야마다 용어의 시기별 출현빈도에 대한 추이는 다른 특징을 지니고 있다는 것을 알 수 있다. '조선'과 같은 용어는 연도별 평균출현빈도는 8건이지만 2011년에는 30건으로 급상승 상위 범위 27건을 상회한다. 용어의 기울기 값은 5.90으로 상대적으로 다른 용어에 비해 높으며 결정계수 값 또한 1에 가까우므로 이는 급상승한 용어라는 것을 알 수 있다. '고향'과 같은 용어는 기울기 값이 -0.30으로써 하락의 추이를 보여준다. 결정계수는 0.03으로 거의 수평에 가까운 형태를 보여준다. 이와 같은 형태는 기울기 값과 결정계수 값이 상당히 작기 때문에 하락을 하더라도 상당히 소폭 하락한다는 것을 의미 한다(<표 5> 참조).

〈표 5〉 인문학 분야 용어의 기울기값과 결정계수

중분류명	용어	기울기값	결정계수	용어유형	비고
한국어외문학	쓰기교육	0.30	0.01	하락후상승형	
한국어외문학	김수영	2.30	0.58	하락후상승형	
역사학	조선	5.90	0.92	상승형	급상승
철학	이성	3.20	0.74	상승형	급상승
한국어외문학	고향	-0.30	0.03	하락형	

- 4) SPSS 19를 통해 산출된 5개의 군집의 형태를 '상승형', '하락 후 상승형', '하락형', '상승후 하락형', '유지형'으로 정의하였다.
- 5) 기울기 값이(절댓값) 클수록 급한 경사를 나타내며 용어 출현의 급상승 또는 급하락인 용어일 가능성이 있다. 급상승과 급하락된 용어에 대한 기준은 편차의 1.5배를 평균에서 뺀 값을 범위로 갖는다. 즉 마지막 년도의 값이 이 범위를 상위로 벗어나면 급상승, 하위로 벗어나면 급하락으로 정의한다.
- 6) 이를 함수로 나타내면 인문학은 $y=0.36x+1.89$, 사회과학은 $y=0.32x+2.06$ 이다.

이처럼 시기별 출현빈도를 분석하여 생명주기적 특성을 도출하고 이를 반영한 출현시기별 패턴에 따라 용어를 구분하여 분석하였다.

5.2 구조적 학술 용어 사전 분석

5.2.1 기초 통계 분석

구조적 학술 용어 사전의 분석은 목적에 따라 다양한 방법을 적용해 볼 수 있다. 앞 장에서 학술논문 주제어를 분석한 것처럼 빈도나, 주제복합성, 네트워크 분석과 같은 방법도 가능하지만, 본 연구의 목적은 용어의 의미관계 형성에 영향을 미치는 요인을 밝히는 것이므로 분석을 수행하는 범위는 종속변인인 의미관계 형성 정도 즉, 관계설정과 관련한 부분으로 제한하도록 한다. 즉 아래 일반 현황에서처럼 총 5,369개의 용어와 이와 관계가 맺어진 용어 노드의 수와 용어 유형의 수에 대한 분석을 수행하였다(〈표 6〉 참조).

5.2.2 용어의 의미관계 노드 수 분석

분야별로 구축된 관계노드수를 살펴보면 다음과 같다. 인문학은 평균 6.7의 관계용어를 가지고 있으며 인문학의 4개의 분석대상 분야를 살펴보면 언어학 5.8, 역사학 7.6, 철학, 9.3, 한국어와 문학이 7개의 평균관계수를 가지고 있

다. 사회과학 분야는 평균 7.2 개의 평균 관계수를 가지고 있는데 분석대상인 경영학은 7.9, 경제학 7.4, 법학은 7.4, 교육학은 6.6 개의 평균 관계수를 가지고 있다. 이처럼 관계수의 평균을 살펴보면 인문학과 사회과학은 큰 차이가 없는 것을 알 수 있다. 또한 분야의 용어 수와 관계용어수의 상관분석을 수행해 본바, 사회과학은 용어의 수와 관계용어수의 상관관계수 값이 0.996, 인문학은 상관관계수 값이 0.991로 사회과학과 인문학 모두 관계용어수와의 관계가 상당히 높다. 하지만 추세를 통해 이를 살펴보면 사회과학의 설명력(R^2)은 99.6%, 인문학의 설명력(R^2)은 98.2%이다. 또한 사회과학 분야는 인문학에 비해 급한 기울기를 가지고 있는 것을 볼 수 있다. 즉 인문학, 사회과학 분야 모두 분야의 용어의 수가 많아질수록 관계용어수도 많아짐을 알 수 있지만 사회과학이 인문학에 비해 용어의 수가 많아질수록 관계용어수의 증가가 더 큰 것을 살펴볼 수 있다.

5.2.3 용어의 의미관계 유형 수 분석

구조적 학술 용어 사전에서는 용어 간의 관계를 의미 속성을 부여하여 연결한다. 구축된 용어의 의미관계 유형은 단 1개만 사용된 유형이 있는 반면에 RT 관계처럼 17,257(48.5%)개로 상당히 많이 사용되는 유형이 있는 것을

〈표 6〉 구축용어 일반현황

용어수	동의어 구축		연구분야 분류 설정수	개념범주 설정		관계 설정수
	구분	용어수		범주설정	개념속성 구축	
5,369	한자	4,560	5,676	5,418	4,513	35,595
	외국어	3,601				
	이칭	2,527				
	음차표기어	41				

알 수 있다. 관계 용어의 수와 관계 유형의 수에 대한 상관관계를 살펴보면 상관계수는 0.628로 비교적 높은 상관관계가 있음을 알 수 있다. 즉 관계 유형수가 많은 경우에 대체로 관계 용어 수도 많다는 것을 의미 한다.

6. 가설 검증

본 연구의 가설 검증을 하기 위해 다음과 같은 통계 방법을 이용하였다.

1. 용어의 의미관계 네트워크 구조에서 용어의 주제복합성이 네트워크 중심성 정도에 미치는 영향관계 분석을 위해서 단순 회귀분석을 이용하였다.
2. 용어의 의미관계 네트워크 구조에서 용어의 위치가 용어의 의미관계 형성정도에 미치는 영향관계분석을 위해 단순회귀분석을 이용하였다.
3. 용어의 출현빈도가 용어의 의미관계 형성 정도에 미치는 영향관계분석을 위해서 단순회귀분석과 일원변량분석을 이용하였다.
4. 네트워크 중심성, 용어의 출현빈도, 구조적 공백성이 용어의 의미관계 형성정도에 미치는 영향관계 분석을 위해 다중 회귀 분석을 이용하였다.

용어의 의미관계 네트워크 구조에서 용어의 주제복합성이 네트워크 중심성에 미치는 영향에 대한 가설 검증 결과를 요약해보면 용어의 브릴로엔 지수는 근접 중심성 정도에만 영향력을 미치고 있으며 연결정도 중심성, 매개 중심성, 위세 중심성에는 영향을 미치고 있지 않다. 하지만 브릴로엔 지수가 근접 중심성에 영향을 주고 있음에 불구하고 모형의 설명력은 낮았다 (<표 7> 참조).

다음으로 용어의 분야 특정 용어 지수는 연결정도 중심성에는 영향력을 미치지 않고 있으며 근접 중심성, 매개 중심성, 위세 중심성에는 영향력을 미치고 있는 것으로 나타났다. 하지만 용어의 분야 특정 용어 지수가 근접 중심성, 매개 중심성, 위세 중심성에 영향을 주고 있음에도 불구하고 모형의 설명력은 낮았다.

용어의 의미관계 네트워크 구조에서 용어의 연결정도 중심성이 의미관계 형성정도에 미치는 영향에 대한 가설 검증 결과를 요약해보면 의미적 연결 관계 수에 영향을 미치는 요인은 연결정도 중심성, 근접 중심성, 매개 중심성, 위세 중심성으로 나타났으며 이 중 매개 중심성은 결정계수(R^2)가 .210로 21%의 설명력을 지니 다른 요인에 비해 높은 설명력(연결정도 중심성의 설명력 11.1%, 근접 중심성의 설명력 2.8%, 위세 중심성의 설명력 2.7%)을 지니고 있음을

<표 7> 용어의 브릴로엔 지수와 근접 중심성 지수와의 관계

모형	비 표준화 계수		표준화 계수	t	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	13.549	.059	-	230.499	13.395 (.000)	.016	-	-
브릴로엔지수	-.331	.090	0.126	3.660			1.000	1.000

종속변수: 근접 중심성 지수

Durbin-Watson = 1.586

* p< .05, ** p< .01, *** p< .001

알 수 있다(〈표 8〉 참조).

다음으로 의미적 연결 유형수에 영향을 미치는 요인은 연결정도 중심성과, 근접 중심성, 매개 중심성이었으며 위세 중심성은 영향을 미치지 않는 것으로 나타났다.

용어의 출현빈도가 의미관계 형성정도에 미치는 영향에 대한 가설 검증 결과를 요약해보면 용어의 출현빈도는 의미관계의 노드의 수와 유형의 수에 모두 영향을 미치고 있었다(〈표 9〉, 〈표 10〉 참조).

용어의 출현 패턴 구분에 따라서는 상승형 패턴의 집단이 유지형 패턴의 집단이나 하락형

패턴의 집단에 비해 의미관계 노드의 수와 유형의 수가 더 많은 것으로 나타났다(〈표 11〉, 〈표 12〉 참조).

용어의 의미관계 네트워크 구조에서 용어의 구조적 공백성이 의미관계 형성정도에 미치는 영향에 대한 가설 검증 결과를 요약해보면 의미적 연결 노드 수에는 효과크기와 제약성이 영향을 미치고 있었지만 효과크기의 결정계수 (R^2)가 .192로 19.2%의 설명력을 지녀 제약성의 2.9%에 비해 높은 설명력을 지니고 있었다. 의미적 연결 유형수에도 효과크기와 제약성 모두 영향을 미치고 있었는데 효과크기의 설명력

〈표 8〉 매개 중심성 지수와 의미관계 노드의 수와의 관계

모형	비 표준화 계수		표준화 계수	t	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	6.809	.248	-	27.500	219.605 (.000)	.210	-	-
매개 중심성	7.467	.504	.459	14.819			1.000	1.000

종속변수: 관계 노드의 수
 Durbin-Watson = 2.007
 * p< .05, ** p< .01, *** p< .001

〈표 9〉 용어의 누적 출현빈도와 의미관계 노드 수와의 관계

모형	비 표준화 계수		표준화 계수	t	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	4.872	.334		14.589	171.006 (.000)	.172	-	-
누적출현빈도	.303	.023	.414	13.077			1.000	1.000

종속변수: 관계 노드의 수
 Durbin-Watson = 1.967
 * p< .05, ** p< .01, *** p< .001

〈표 10〉 용어의 누적 출현빈도와 의미관계 유형 수와의 관계

모형	비 표준화 계수		표준화 계수	t	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	3.464	.100		34.620	70.865 (.000)	.079	-	-
누적출현빈도	.059	.007	.281	8.418			1.000	1.000

종속변수: 관계 유형의 수
 Durbin-Watson = 2.049
 * p< .05, ** p< .01, *** p< .001

<표 11> 용어의 시기별 출현 패턴 구분에 따른 용어의 의미적 연결 관계 노드의 일원배치분산분석결과

종속변수	시기별출현패턴	평균	표준편차	F값(p)	Dunnett T3
의미적 연결관계 노드	상승형(a)	7.66	6.765	16.801 (.000)	a>c>b
	유지형(b)	6.43	4.759		
	하락형(c)	6.79	4.963		

* p< .05, ** p< .01, *** p< .001

<표 12> 용어의 시기별 출현 패턴 구분에 따른 용어의 의미적 연결 관계 유형의 일원배치분산분석결과

종속변수	시기별출현패턴	평균	표준편차	F값(p)	Dunnett T3
의미적 연결관계 유형	상승형(a)	3.41	1.82	6.871 (.000)	a>c>b
	유지형(b)	3.17	1.680		
	하락형(c)	3.26	1.710		

* p< .05, ** p< .01, *** p< .001

이 6.2%로 제약성의 1%에 비해 설명력이 더 높았다. 결과적으로 용어의 의미관계 네트워크 관계에서는 효과크기가 제약성에 비해 더 높은 영향력을 지니는 것을 알 수 있다.

용어의 네트워크 중심성, 출현빈도, 구조적 공백성중 어떠한 요인이 용어의 의미관계 형성에 더 많은 영향을 미치는지에 대한 연구 문제를 해결하기 위한 영향관계 검증은 4개의 네트

워크 중심성중 다른 요인에 비해 가장 높은 설명력(21%)을 지니고 있는 매개 중심성, 출현빈도, 구조적 공백성 요인 중 다른 요인에 비해 높은 설명력(19.2%)을 지니고 있는 효과크기를 요인으로 선택한 후 회귀 분석을 통한 가설 검증을 실시하고 유의수준 .05에서 통계적 유의성을 검증하였다. 이를 요약해보면 매개 중심성, 출현빈도, 구조적 공백성의 효과크기는

<표 13> 매개 중심성, 출현빈도, 효과크기가 의미관계 노드 수에 미치는 영향 관계

모형	비 표준화 계수		표준화 계수	t (p)	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	6.175	.456	-	13.534	6.8062 (.000)	.079	-	-
매개 중심성	5.508	1.223	.338	4.504 (.000)			.169	5.906
출현빈도	-.018	.057	-.025	-.318 (.750)			.159	6.271
효과크기	.102	.056	.157	1.833 (.067)			.129	7.735

종속변수: 관계 노드의 수

Durbin-Watson =2.018

* p< .05, ** p< .01, *** p< .001

〈표 14〉 매개 중심성, 출현빈도, 효과크기가 의미관계 유형 수에 미치는 영향 관계

모형	비 표준화 계수		표준화 계수	t (p)	F값 (p)	R ²	공선성 통계량	
	B	표준오차	베타				공차	VIF
(상수)	3.401	.140	-	24.244	23.779 (.000)	.080	-	-
매개 중심성	-.254	.376	-.055	-.675 (.500)			.169	5.906
출현빈도	-.068	.017	.326	3.896 (.000)			.159	6.271
효과크기	.001	.017	.003	.036 (.971)			.129	7.735

중속변수: 관계 유형의 수

Durbin-Watson = 2.052

* p < .05, ** p < .01, *** p < .001

연결관계 노드의 수에 영향을 미치고 있는 것을 알 수 있었다. 이 중 매개 중심성이 가장 큰 영향을 미치고 있으며 효과크기와 출현빈도의 경우 영향을 거의 미치고 있지 못한 것을 알 수 있다. 연결 관계 유형의 수에 영향을 미치고 있는 요인을 살펴보면 출현빈도가 가장 큰 영향을 미치고 있으며 매개 중심성과 구조적 공백성의 효과크기는 영향을 거의 미치지 못하고 있음을 알 수 있다(〈표 13〉, 〈표 14〉 참조).

7. 결론

지금까지 다양한 분야의 전문용어 사전이 구축되면서 용어를 선정하는 프로세스의 주된 방법으로는 용어의 출현빈도가 사용되었다. 하지만 시소러스에 비해 용어 간의 관계성이 더욱 강조되는 구조화된 전문용어 사전을 구축하기 위해서는 기존 용어 선정 방법론인 용어의 출현빈도만으로는 사전의 질을 높이기에는 한계가 있다. 본 연구는 구조적 정의에 의한 학술용어 사전 데이터베이스 구축을 위해 학술용어

의 의미관계 형성에 미치는 요인이 무엇인지를 찾아내고 이러한 요인이 어떠한 영향을 미치고 있는지를 밝히고자 하였다.

본 연구를 위해 의미관계 형성방법론에 관한 이론 및 선행 연구를 통해 용어의 의미관계 형성에 영향을 미칠 것으로 예상되는 요인을 도출하였으며 의미관계 형성 영향관계에 대한 연구모형을 설계하였다. 이어서 학술논문 주제어에 대한 분석을 인문학 4개 분야인 '역사학', '철학', '언어학', '한국어와 문학', 사회과학 4개 분야인 '경영학', '경제학', '법학', '교육학' 총 8개 학문 분야로 나누어 살펴보고 분야에 속한 용어들이 네트워크 구조에서 어떠한 위치를 차지하고 있는지를 살펴보았다. 그리고 이를 통해 용어의 주제복합성, 용어의 네트워크 상 위치, 용어의 출현빈도가 용어 간의 의미관계 형성에 미치는 영향을 검증하였다.

용어의 주제복합성이 네트워크 중심성에 미치는 영향에 대한 가설 검증 결과, 브릴로엔 지수는 근접 중심성에만 영향력을 미치고 있으며 연결정도 중심성, 매개 중심성, 위세 중심성은 영향을 미치고 있지 않은 것으로 나타났다. 용

어의 분야 특정 용어 지수는 연결정도 중심성에는 영향력을 미치지 않고 있으며 근접 중심성, 매개 중심성, 위세 중심성에는 영향을 미치고 있는 것으로 나타났다.

용어의 의미관계 네트워크 구조에서 용어의 위치가 용어의 의미관계 형성에 미치는 영향에 대한 가설 검증을 수행한 결과 용어의 의미적 연결 관계 수에 영향을 미치는 요인은 연결정도 중심성, 근접 중심성, 매개 중심성, 위세 중심성인 것으로 나타났다. 의미적 연결 유형수에 영향을 미치는 요인은 연결정도 중심성과, 근접 중심성, 매개 중심성이었으며 위세 중심성은 영향을 미치지 않는 것으로 나타났다.

용어의 누적 출현빈도가 용어의 의미관계 형성에 미치는 영향에 대한 가설 검증을 수행한 결과 의미관계의 노드와 의미관계 유형의 수에 모두 영향을 미치고 있는 것으로 나타났다. 용어의 출현 패턴 구분에 따라서는 상승형패턴의 집단이 유지형패턴의 집단이나 하락형패턴의 집단에 비해 의미관계 노드의 수와 유형의 수가 더 많은 것으로 나타났다.

용어의 구조적 공백성이 용어의 의미관계 형성 정도에 미치는 영향에 대한 결과 의미적 연결 노드 수에는 효과크기와 제약성이 영향을 미치고 있었다. 의미적 연결 유형수 또한 효과크기와 제약성 모두 영향을 미치고 있었으며 결과적으로 용어의 의미관계 네트워크 관계에서는 효과크기가 제약성에 비해 더 높은 영향력을 지니는 것으로 나타났다.

마지막으로 용어의 매개 중심성, 출현빈도, 구조적 공백성의 효과크기가 의미적 연결 관계 형성 정도에 미치는 영향에 대한 가설 검증 결과를 살펴보면 매개 중심성, 출현빈도, 구조적

공백성의 효과크기는 연결관계 노드의 수에 영향을 미치고 있는 것으로 나타났다. 이 중 매개 중심성이 가장 큰 영향을 미치고 있으며 다음으로 효과크기의 영향이 크고 출현빈도의 경우 영향을 거의 미치고 있지 못한 것으로 나타났다. 연결 관계 유형의 수에 영향을 미치는 요인을 살펴보면 출현빈도가 가장 큰 영향을 미치고 있으며, 매개 중심성, 효과크기는 영향을 거의 미치지 못한 것으로 나타났다.

가설 검증을 통해 구조적 학술용어사전의 구축에 있어 의미관계 형성 정도에 주요한 영향을 미치는 요인으로 매개 중심성, 출현빈도, 구조적 공백성의 효과크기라는 것을 알 수 있었다. 그리고 용어의 중요성은 일반적으로 알려진 출현빈도를 통한 방법 외에도 연결정도 중심성, 근접 중심성, 매개 중심성, 위세 중심성과 같은 측정방법에 따라서도 판단할 수 있음을 확인하였다. 또한 주제복합성은 직접적으로 의미관계 형성 정도에 영향을 미치지 않지만 용어의 근접 중심성에 영향을 미치기 때문에 크게 4가지의 요인을 고려하여 용어를 선정할 경우 의미관계 형성 정도는 높아질 수 있는 것으로 파악되었다.

본 연구의 결과는 지금까지 용어를 선정하는 프로세스의 주된 방법론인 용어의 출현빈도를 활용하는 방법 이외에도 용어 네트워크상에서의 용어의 위치나 주제복합성 같은 방법론을 적용하여 용어를 선정할 수 있다는 것을 보여준다. 특히 용어 간의 관계성이 더욱 강조되는 구조화된 전문용어 사전 구축 시에는 용어 네트워크 구조에서의 매개 중심성 값을 활용하여 용어를 선정하는 것이 출현빈도를 활용한 방법보다 용어의 의미관계 형성 정도에 더 많은 영

향을 미칠 수 있음을 보여준다. 따라서 전문용어 사전의 구축할 때 용어의 네트워크에서의 매개 중심성, 출현빈도, 구조적 공백성의 효과 크기, 용어의 주제복합성을 면밀히 판단하여 다각도로 용어를 선정할 경우 전문용어 사전의 질적인 향상과 완성도가 높아질 것을 기대할 수 있다.

본 연구는 학술 용어의 선정 분야를 전분야가 아닌 인문학과 사회과학 분야로 한정하였으며 구조적 학술용어사전이라는 특수한 대상에

적용하였으므로 향후 주제 분야의 확장과 구조적 관계를 지닌 다른 사전의 구축에도 적용을 하는 등 다양하게 적용해보는 실험적 연구가 필요하다. 또한 본 연구에서 제시된 용어 선정 방식은 의미관계형성정도에 미치는 영향만을 통해 검증한 한계가 있으므로 연구의 목적과 범위에 따라 의미관계형성정도가 아닌 다른 변인을 설정하고 용어 선정 방식의 적합성과 효율성을 측정할 수 있는 다양한 방법들을 적용한 후속 연구가 이루어질 필요가 있다.

참 고 문 헌

- [1] 강성희, 박승수. 1995. 의미 구조를 갖는 시소러스의 지식 표현에 관한 연구. 『한국정보과학회 학술 발표논문집』, 22(1): 183-186.
- [2] 강정미. 1999. 전문용어사전 표제어 기술형식에 대한 연구. 『韓國情報管理學會 學術大會論文集』, 6: 39-42.
- [3] 고영만. 2007. 의미 호환을 위한 메타데이터 매핑 연구. 『정보관리학회지』, 24(4): 223-238.
- [4] 고영만, 송민선, 김비연, 민혜령. 2013. 인문학 및 사회과학 분야 국내 학술논문의 저자키워드 출현 빈도와 피인용횟수와의 상관관계 연구. 『정보관리학회지』, 30(2): 227-243.
- [5] 김원중. 2003. 『시소러스를 이용한 역사 정보검색 시스템의 구축 및 효율성 평가 연구』. 석사학위논문, 성균관대학교 대학원 문헌정보학과.
- [6] 김은영. 2010. 『디자인사전의 편찬방향 연구』. 석사학위논문, 홍익대학교 대학원 시각디자인전공.
- [7] 김일환, 강범모, 김흥규. 2012. 공기어를 이용한 사회 문제에 대한 관심의 추이 분석 - 2000~2009년 신문기사를 중심으로 -. 『인문학연구』, 21: 175-202.
- [8] 김태수. 2001. 용어 정의를 도입한 시소러스 개발 연구. 『情報管理學會誌』, 18(2): 231-254.
- [9] 김관준, 이재운. 2012. 디스크립터 자동 할당을 위한 저자키워드의 재분류에 관한 실험적 연구. 『情報管理學會誌』, 29(2): 225-246.
- [10] 김현수. 2010. 『연구자 네트워크의 구조적 위치가 연구성과에 미치는 영향에 관한 연구: 국내 나노테크놀로지 분야를 중심으로』. 석사학위논문, 성균관대학교 대학원 문헌정보학과.
- [11] 김희정. 2011. 네트워크 분석을 기반으로 한 웹 이카이빙 주제영역 연구. 『한국비블리아학회지』, 22(2): 235-248.

- [12] 남영준, 이두영. 1994. 자연어처리를 이용한 시소러스 자동생성. 『韓國情報管理學會 學術大會論文集』, 1: 89-92.
- [13] 박옥남. 2011. 정보조직 지식구조에 대한 연구: 2000년~2011년 학술논문을 중심으로. 『한국비블리아학회지』, 22(3): 247-267.
- [14] 박정오, 황도삼. 2000. 전문용어 추출시스템. 『한국정보과학회 학술발표논문집』, 27(1B): 381-383.
- [15] 서선경, 정은경. 2013. 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. 『한국비블리아학회지』, 24(1): 207-228.
- [16] 송유화. 2012. 『Inspec 시소러스를 이용한 실질적인 연관관계 유형에 관한 연구』. 석사학위논문, 과학기술연합대학원대학교 대학원 응용정보학과.
- [17] 양창진. 2010. 학술 논문의 주제어 표기 및 활용 방안 연구. 『인문콘텐츠』, 19: 395-416.
- [18] 유영준. 2005. 온톨로지의 개념간 관계 설정을 위한 AGROVOC 시소러스의 분석에 관한 연구. 『情報管理學會誌』, 22(1): 125-144.
- [19] 이민우. 2011. '명사+명사' 합성어 구성요소의 의미관계 연구. 『한국어 의미학』, 34: 235-257.
- [20] 이수경, 정상원, 김흥기, 염영희. 2011. 한국 간호학 연구주제의 사회 연결망 분석. 『대한간호학회지』, 41(5): 623-632.
- [21] 이수진. 2012. 『핵심어 분석을 통한 학술 핵심어의 어휘 특성 연구』. 석사학위논문, 경북대학교 대학원 국어국문학과.
- [22] 장정우, 최경호. 2012. 언어네트워크분석을 이용한 통계법 내용분석. 『통계연구』, 17(2): 53-66.
- [23] 정성훈. 2012. 『이슈 프레임 분석을 위한 전통적 내용분석과 언어네트워크분석 방법 비교연구: 부안 원전수거물 관리시설에 대한 언론보도를 중심으로』. 석사학위논문, 한양대학교 대학원 광고홍보학과.
- [24] 조개나. 2012. 『한국어 합성명사의 형태·의미적 특성 연구』. 석사학위논문, 강원대학교 대학원 국어국문학과.
- [25] 조은경. 2001. 『전문용어의 어휘형태적 특성 연구』. 석사학위논문, 연세대학교 대학원 국어정보학협동과정.
- [26] 조재인. 2011. 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석. 『정보관리학회지』, 28(4): 65-83.
- [27] 조현양, 남영준. 2004 시소러스와 온톨로지의 상호 호환성에 관한 연구. 『情報管理學會誌』, 21(4): 27-47.
- [28] 주홍근. 2010. 『국어 인체 명칭 핵 명사의 조어법 특성 연구』. 석사학위논문, 강원대학교 대학원 국어국문학과.
- [29] 최영출, 박수정. 2011. 한국행정학의 연구경향 분석: 네트워크 텍스트 분석방법의 적용. 『韓國行政學報』, 45(1): 123-139.

[30] Burt, R. 1992. *Structural Holes*. Cambridge: Harvard University Press.

• 국문 참고자료의 영어 표기

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

- [1] Kang, Sung Hee and Park, Seung Soo. 1995. "Knowledge Representation for the Thesaurus with Semantics," *Proceedings of Journal of Computing Science and Engineering*, 22(1): 183-186.
- [2] Kang, Jung-mi. 1999. "A Study on the Entry Description Format of Glossaries," *Proceedings of Journal of the Korean Society for Information Management*, 6: 39-42.
- [3] Ko, Young Man. 2007. "A Study on Metadata Mapping for Semantic Interoperability," *Journal of the Korean Society for Information Management*, 24(4): 223-238.
- [4] Ko, Young Man, Song, Min-Sun, Kim, Bee-Yeon and Min, Hye-Ryoung. 2013. "A Study on the Correlation between the Appearance Frequency of Author Keyword and the Number of Citation in the Humanities and Social Science Journal Articles of the Korea Citation Index (KCI)," *Journal of the Korean Society for Information Management*, 30(2): 227-243.
- [5] Kim, Wan Jong. 2003. *Research on Construction and Merit Rating of Retrieval System for the Historical Information Using Thesaurus*. M.A. thesis, Sungkyunkwan University.
- [6] Kim, Eun-young. 2010. *Prerequisite for the creation of a design dictionary: An analysis of current and past data compilation methods*. M.A. thesis, Sungkyunkwan University.
- [7] Kim, Ii Hwan, Kang, Beom Mo and Kim, Heung Gyu. 2012. "The Analysis of Social Problems based on keywords and their associated words from a large-scale newspaper corpus - focused on the newspaper issued from the year 2000 to 2009 -." *The Journal of humanities*, 21: 175-202.
- [8] Kim, Tae-Soo. 2001. "A Study on the Development of Thesaurus Using Terminological Definitions," *Journal of the Korean Society for Information Management*, 18(2): 231-254.
- [9] Kim, Pan Jun and Lee, Jae Yun. 2012. "A Study on the Reclassification of Author Keywords for Automatic Assignment of Descriptors," *Journal of the Korean Society for Information Management*, 29(2): 225-246.
- [10] Kim, Hyun soo. 2010. *Influence of researcher network structure on research performance: applied to Korean nano technology researchers*. M.A. thesis, Sungkyunkwan University.
- [11] Kim, Heejung. 2011. "A Study on Web Archiving Subject Analysis Based on Network Analysis," *Journal of the Korean Biblia Society for Library and Information Science*, 22(2): 235-248.
- [12] Nam, young-joon and Lee, Too-young. 1994. "Automatic construction of thesaurus using natural language processing," *Proceedings of Journal of the Korean Society for Information*

- Management*, 1: 89-92.
- [13] 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.
- [14] Park, Jung-Oh and Hwang, Do-sam. 2000. "A terminology extraction system." *Proceedings of Journal of KISS*, 27(1B): 381-383.
- [15] Seo, SunKyung and Chung, EunKyung. 2013. "Domain Analysis on the Field of Open Access by Co-Word Analysis." *Journal of the Korean Biblia Society for Library and Information Science*, 24(1): 207-228.
- [16] Song, yoohwa. 2012. *Study on the Associative Relationships using Inspec Thesaurus*. M.A. thesis, University of Science and Technology.
- [17] Yang, Chang-Ji. 2010. "Study on Keywords and Their Use of Academic Theses - Focused on Database Development and Information Link." *Korea Humanities Content Society*, 19: 395-416.
- [18] 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.
- [19] Lee, minwoo. 2011. "A Study on Semantic Relations between the Components of 'Noun + Noun' Compounds." *Korean Semantics*, 34: 235-257.
- [20] Lee, Soo-Kyoung, Jeong, Senator, Kim, Hong-Gee and Yom, Young-Hee. 2011. "A Social Network Analysis of Research Topics in Korean Nursing Science." *J Korean Acad Nurs*, 41(5): 623-632.
- [21] Lee, Su-jin. 2012. *A Keyword Analysis of Academic Vocabulary in Korean: Focused on academic keywords in Korean literature and linguistics*. M.A. thesis, Kyungpook National University.
- [22] Jang, Jeong-woo and Choi, kyoung-Ho. 2012. "Statics Act Content Analysis using Semantic Network Analysis." *Journal of the Korean Official Statistics*, 17(2): 53-66.
- [23] Jung, Sung Hoon, 2012. *Traditional content analysis and sementic network analysis comparative research In issue framing analysis*. M.A. thesis, Hanyang University.
- [24] Zhao, Kai-Na. 2012. *A Study on the Morphological and Semantic Characteristics of Korean Compound Nouns*. M.A. thesis, Kangwon National University.
- [25] Jo, EunKyoung. 2001. *A Study on Lexico-Morphological characteristics of Terminology*. M.A. thesis, Yonsei University.
- [26] Cho, Jane. 2011. "A Study for Research Area of Library and Information Science by

- Network Text Analysis.” *Journal of the Korean Society for Information Management*, 28(4): 65-83.
- [27] 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.
- [28] Zhou, Hong-gun. 2010. *A study on Characteristics of word formation for core nouns of body part names*. M.A. thesis, Kangwon National University.
- [29] Choi, youngchul and Park, soojung. 2011. “A analysis of Korean Public Administration Research Trend.” *Korean Public Administration Review*, 45(1): 123-139.