

패킷 분석 기법을 활용한 건설 시소러스 구축 방안 에 관한 연구*

A Study on Building Method of the Construction Industry Thesaurus Using Facet Analysis Method

홍 기 철(Ki-Churl Hong)**

〈 목 차 〉	
I. 서론	1. 개발 원칙
1. 연구의 목적 및 배경	2. 개발 절차
2. 연구의 범위, 방법 및 절차	3. 절차별 개발 방안
II. 이론적 배경	4. 패킷의 적용방안
1. 패킷 분석	IV. 패킷 분석 기반 건설 시소러스 구축 방안
2. 주제관련성과 범주관련성	1. 시소러스의 설계 방안
3. 구분의 특성	2. 구축 절차
4. 용어와 개념	3. 절차별 구축 방안
5. 선행 연구	4. 구축 사례
III. 건설 분야 패킷 유형 개발	V. 결론

초 록

전통적으로 건설 산업 분야는 패킷 분석 기법을 활용한 대표적인 영역 중의 하나이다. 국제적으로 분류체계인 『건설정보 통합분류법(Uniclass)』이나 시소러스인 『건설산업시소러스(Construction Industry Thesaurus, CIT)』 등에서 패킷 분석 기법을 적용하고 있다. 국내의 경우, 분류체계인 『건설정보분류체계』 등에 활용되고 있으나 시소러스에서의 패킷 분석 기법의 연구나 구축 사례는 찾아보기 힘들다. 본 연구의 목적은 패킷 분석 기법을 활용한 건설 분야의 정보검색을 위한 시소러스의 구축을 상정하여 패킷 유형을 확립하고, 패킷 분석 기법에 따른 건설 분야 시소러스의 구축 방안을 제시하는 데에 있다. 기본 패킷으로 최상위 10개의 패킷(주체 및 수동체, 인공물, 추상물, 재료, 부품/구성요소, 공중, 속성, 매체, 프로세스, 공간, 시간)과 하위 20개의 패킷을 설립하고, 국제표준(ISO 25964-1)이 제시하고 있는 시소러스 구축 절차에 따른 구축 방안을 제시하였다. 본 연구의 결과는 패킷 기반 시소러스의 적용뿐만 아니라 향후 택사노미나 온톨로지 등에서의 재활용 및 건설 분야의 분류체계 등과의 상호운용성에 활용이 가능할 것으로 기대하고 있다.

키워드: 패킷 분석, 기본 패킷, 건설 시소러스, 시소러스 구축 방안, ISO 25964-1

ABSTRACT

Traditionally, the construction industry field is one of the typical fields using facet analysis method. Internationally the facet analysis method is applied to the classification scheme 「Uniclass」 or thesaurus 「Construction Industry Thesaurus(CIT)」. In the case of Korea, the facet analysis method is being used in classification scheme 「Construction Industry Classification System」 but it is difficult to find studies on facet analysis method in thesaurus or building case study. This study aims to establish facet types by introducing building of thesaurus for information retrieval in construction field using facet analysis method and to suggest building plan of thesaurus in construction field using facet analysis method. In this paper, We establish the 11 top facets(agent/patient, artifacts, abstract, material, parts/component, works, attribute, media, process, space, time) and the 13 subfacet as rudimental facets and suggest building plan according to thesaurus building procedure which is suggested by International Standard(ISO 25964-1). The result of this study is expected to apply to facet based thesaurus. And it is also expected to reuse in Taxonomy or Ontology etc. and to use in interoperability with the classification scheme of Construction filed.

Keywords: Facet analysis, Fundamental facets, Construction Industry Thesaurus, Thesaurus construction method, ISO 25964-1

* 이 논문은 한국건설기술연구원의 “해외 거점별 건설엔지니어링 정보시스템 개발” 연구의 일환으로 진행되었음.

** 한국건설기술연구원 객원연구원(hongkc3965@naver.com)

•논문접수: 2017년 2월 20일 •최초심사: 2017년 2월 28일 •게재확정: 2017년 3월 22일

•한국도서관정보학회지 48(1), 345-371, 2017. [http://dx.doi.org/10.16981/kliss.48.201703.345]

I. 서론

1. 연구의 목적 및 배경

패킷 분석은 1960년대에 Thesaurofacet으로 시작하여 온라인상에서 가장 널리 사용되는 시소러스 중 하나인 『예술과 건축 시소러스(Art and Architecture Thesaurus, AAT)』를 포함한 계층구조를 구성하는 패킷 분석을 사용하는 시소러스의 구축에서 인기 있는 선택이다(Ryan 2014, 20). 패킷 분석은 “시소러스의 구축을 위한 최상의 출발점이 몇 가지 종류의 분류로 확고하게 확립된 것”이며, “지식조직화 시스템의 어떤 유형의 구축의 기초가 되어야 하는 중요한 기법”으로 정보검색을 위한 시소러스의 구축 국제규격인 ISO 25964에 채택되었다(Ryan 2014, 20).

최근, 주제전거에 대한 기능상의 요건(Function on Requirement for Subject Authority Data, FRSAD)에 따르면 패킷은 하나의 개념체계를 세우는 것은 불가능하며(IFLA 2010, 17), 패킷의 선택은 주제 분야에 따라 달라질 수 있다(ISO 25964-1 2011, 68).

시소러스가 기타의 제어 어휘와 크게 다른 점은 구분의 원리가 동일한 상하위의 계층구조를 가지고 있는 점이다. 시소러스를 구축함에 있어서 패킷 분석은 시소러스의 계층구조를 생성하는 데에 동일한 범주에 속하는 개념에 대해서만 타당하기 때문에 유용하다(ISO-25964-1 2011, 68).

특히, 전통적으로 건축 산업 분야는 패킷분류표의 주요 이용자였으며, 이 분야의 정보를 처리하기 위해 다양한 패킷분류표가 개발되었다(Hunter 2015, 46). 대표적인 예로는 『브리슈 건축분류법』, 『CI/SfB분류법』 및 『건설정보통합분류법(Uniclass)』 등이 있으며, 국내에서는 『건설정보분류체계』가 있다. 또한 패킷 분석 기법을 적용한 정보검색용 시소러스도 1971년에 공표되어 현재는 『Building Services Thesaurus』로 개칭된 『Construction Industry Thesaurus(CIT)』(이하 『건설산업시소러스(CIT)』)가 있다.

하지만, 국내에서 건설 분야의 패킷 분석 기법을 이용한 시소러스의 구축 및 활용에 대한 연구는 거의 찾을 수 없다. 또한, 국내에서는 건설 분야에 대해 구축된 시소러스가 없고, 다만 시소러스의 핵심이라고 할 수 있는 개념의 상하관계가 확립되지 않은 『건설용어 동의어 및 관련어집』(한국건설기술연구원 1996)이 있을 뿐이다.

패킷 유형을 개발하고, 패킷 분석 기법을 활용하면, 시소러스의 통합(Thesauri Integration)이나 분류표, 주제명표, 이름전거데이터, 동의어사전(Synonym ring)¹⁾ 및 온톨로지 등 타

1) ANSI/NISO Z39.1(NISO 2005)에서처럼 제어어휘로 구축되었을 경우에 상호운용성을 고려할 수 있다.

어휘와의 상호운용성을 고려함에 있어서도 패킷 분석 기법을 활용할 수 있다. 특히, 국내의 경우 패킷 기반 시소러스가 구축되면, 『건설정보분류체계』와 시소러스(후조합시스템)-분류체계(전조합시스템)간의 매핑에 의한 상호운용성을 고려할 수 있을 것이다.

이에 본 연구는 건설산업 분야의 패킷 분석 기법을 이용한 시소러스 구축을 위한 기본 패킷(fundamental facet)을 확립하고, 국제표준(ISO 25964-1 2011)의 시소러스 구축 절차에 따른 건설 분야 시소러스의 구축 방안을 제시하고자 한다.

2. 연구의 범위, 방법 및 절차

본 연구는 패킷 분석 기법을 적용한 건설 분야의 정보관리를 위한 시소러스 구축을 위한 기본 패킷을 확립하고, 그것을 기반으로 하는 건설 분야 시소러스 구축 방안을 제시하는 것을 연구 목표로 삼았다.

기본 패킷 및 패킷 분석에 있어서 주제관련성(aboutness)과 범주관련성(isness)을 고려할 수 있다. 본 연구에서 범주관련성은 고려하지 않았다. 또한, 시소러스의 구축 방안 에 있어서는 구축 이후의 서문의 작성이나 유지와 보수 등에 대해서는 간략하게 설명하는데 그쳤다.

본 연구는 문헌연구에 의해 패킷과 패킷 분석을 파악하였으며, 선행연구에서 패킷 유형의 개발, 패킷의 구조화, 패킷의 적용 사례 등을 분석하였다. 한편 사례 분석은 건설 분야의 정보관리를 위한 패킷을 개발함에 있어서 기존의 기본 패킷은 영역을 한정하지 않았다. 여기서는 전 주제분야 영역과 과학기술 분야의 전체 영역, 그리고 건설 분야에 한정된 영역의 기본 패킷을 조사·분석하고, 이를 바탕으로 패킷 분석을 통해, 하향식 방식으로 임시 기본 패킷(카테고리)을 설정하였다. 본 연구에서는 『건설정보분류체계』(건설교통부 2006)의 용어 및 개념을 분석하여 임시 기본 패킷을 보완하여 건설 분야 기본 패킷을 확립하였다.

시소러스의 구축 절차는 국제표준(ISO 25964-1 2011)과 아일랜드 시소러스구축 지침(Ryan 2014)의 문헌분석을 통하여 제시하였고, 해당 절차에 따라 패킷 분석 기반의 건설 분야 정보관리를 위한 시소러스 구축 방안을 도출하고 시험용 시소러스를 구축하였다.

Ⅱ. 이론적 배경

1. 패킷 분석

패킷이란 주제 분야를 구성하는 주요 속성, 관점, 혹은 특성을 뜻하며, 기본 주제를 일련의 공통적인 특성에 따라 여러 개의 기본 범주로 구분할 때 이 범주들을 기본 패킷이라 한다(최

윤경, 정연경 2015, 183). 기본 패킷은 카이저(J. Kaiser)에 의해 최초로 명명되었지만, 분석-합성 분류의 고안자는 랑가나단(R. Ranganathan)으로 그가 패킷 분석을 개발하고 대중화하여 적용한 것이 콜론 분류법이다. 그는 콜론 분류법에서 패킷으로 다섯 개의 기본 카테고리를 열거한 것이 성격(personality), 물질(matter), 에너지(energy), 공간(space) 그리고 시간(time)으로 이것을 PMEST라고 한다(Aitchison et al. 2000, 70).

패킷 분석은 시소러스에서 계층 관계에 대한 규칙을 따르는 계층구조를 생성하는 데 유용하다. 이러한 계층관계는 동일한 범주에 속하는 개념에 대해서만 타당하기 때문이다. 패킷의 선택은 주제 분야에 따라 달라질 수 있지만, 최상위 수준에서는 대상(object), 재료, 주제, 행위, 장소, 시간 등과 같은 기본 범주(fundamental categories)를 사용하는 것이 일반적이다(ISO 25964-1 2011, 68).

기본 패킷은 필요한 수준까지 하위 패킷으로 분석될 수 있다. 예를 들어, 행위(action)는 “숙성” 또는 “열화”와 같은 자동사적인 프로세스(과정)와 “절단” 또는 “복구”와 타동사적인 오퍼레이션(작용)으로 하위 구분할 수 있다. 여기서 행위는 “행위자”가 어떤 대상에 영향을 받지 않을 때 자동사라고 하고, 반면에 타동사적 행위의 행위자는 어떤 대상에서 영향을 받는다. 예를 들어 “사과”는 내부 프로세스에 의해 익지만, “사과 절단”에 있어서 칼은 외부 대상으로 영향을 준다(ISO 25964-1 2011, 68). 패킷 분석은 “패킷으로 그룹화된 구성요소 개념들로 주제의 영역을 분석, 그리고 구분의 특성에 의해 하위 개념들로 개념의 세분화”로서 정의하고 있다(Ryan 2014, 21, ISO 25964-1 2011, 68).

기본 패킷은 특정의 주제 분야에 강제되어야 하는 것은 아니다. 그러나 기본 패킷은 분류별 주제를 잘 배열하는 것에 편리한 틀을 제공하고, 또한 간과해서는 안 되는 패킷에 대한 안내가 되는 것이다. 그렇지만, 개개의 분야에는 각각에 특수한 문제가 있고, 그 분야의 검색시스템의 이용자에 있어서 가장 유효한 방법으로 구분되어야 한다(Lancaster 1972, 39).

기본 패킷을 가진 시소러스는 먼저 주제 분야나 학문 분야로 나눈 시소러스에 의해 커버되는 지식의 영역임에도 불구하고, 주요구분으로서 먼저 기본 패킷으로 나누는 것도 하나의 대안이다. 이런 사례로 초기의 것으로 『건설산업시소러스(CIT)』가 있으며 최근의 사례로는 『예술 및 건축 시소러스(AAT)』가 있다(Aitchison et al. 2000, 73).

여기서 패킷 내의 계층구조에서 용어들은 모두 동일 기본 패킷이다. 이것은 모든 상호 용어 관계는 전체-부분 계층 관계이며, 연관관계는 아님을 보증한다. 상이한 계층구조에서 용어들 사이의 연관관계는 완전한 계층구조와 동반하는 자모순 표시에서만 보여준다(Aitchison et al. 2000, 75). 패킷은 상위 온톨로지와 매우 유사한 면을 가지고 있다. 다만, 온톨로지에서는 클래스가 개체(individual)를 구성요소로 가지지만, 시소러스에서는 주제를 표현하기 때문에 그 하위의 전개는 사뭇 다르다. 기존의 국제규격(ISO 2788)에서는 언급이 없었으나,

새로운 국제규격(ISO 25964)에서는 기술하고 있다.

2. 주제관련성과 범주관련성

FRSAR(Functional Requirements for Subject Authority Records)는 일부 제어어휘집이 주제뿐만 아니라 형식이나 장르, 이용 대상과 같이 저작의 다양한 측면을 표현하는 용어를 제공하고 있다. 그러한 측면은 매우 중요하며, 다수의 이용자 질의에서 나타난다. 즉, 저작이 무엇에 관한 것인지 보다는 범주관련성(isness) 또는 형식이나 장르(예. 소설, 희곡, 시, 수필, 전기, 교향곡, 협주곡, 소나타, 지도, 도면, 그림, 사진 등)에 기초하여 해당 저작이 어느 부류에 속하는지를 기술하는 것이지만 엄격하게 보면 이것들은 패킷이 아니다(Ryan 2014, 22, IFLA 2010, 10). 따라서 시소러스 내에서의 범주관련성은 그다지 사용되지 않는다. 그러나 인터넷 등 온라인 표시에서는 거의 필수가 되고 있다.

3. 구분의 특성

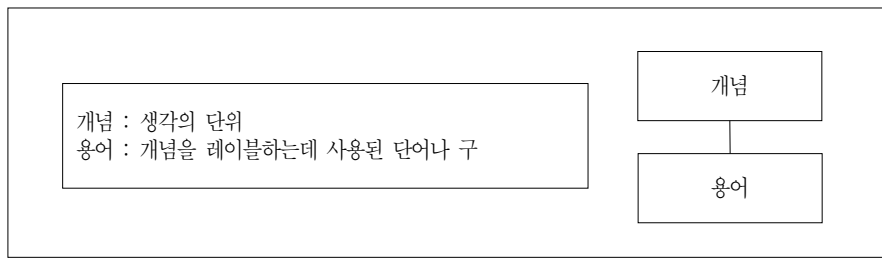
구분의 특성은 해당 속성의 다른 값을 갖는 하위개념의 동위개념(array)으로 세분되어질 수 있는 개념의 속성을 말한다. 여기서 “패킷 분석은 패킷으로 그룹화된 구성요소 개념들로 주제의 영역을 분석, 그리고 구분의 특성에 의해 좁은 개념의 세분화”로 정의된다(ISO 25964-1 2011, 2). 하나의 고려사항은 다중 계층 구조를 가지는 경우, 구분의 특성에 의한 하위 구분은 엄격하게 적용할 수 없다. 예를 들어, 두개골의 경우, 머리와 뼈의 두 상위개념을 가질 수 있다. 전자의 관계는 전체-부분 관계이고, 후자의 관계는 속종관계이기 때문에 논리적으로 다른 관계에 기초하고 있다(ISO 25964-1 2011, 62)

4. 용어와 개념

시소러스는 용어에 의해 표현된 개념들의 통제되고 구조화된 어휘이다(ISO 25964-1 2011, 12). ISO 25964-1은 생각의 단위인 개념들, 그리고 우리가 그들을 표현하는 데 사용하는 용어들 차이를 명시적으로 나타내는 최초의 시소러스 표준이다(Ryan 2014, 12). 여기서 개념들은 그것들을 표현하는 데에 사용된 용어들에 독립적으로 추상적 개체로서 마음속에 존재한다. 예를 들어, “어린이”는 매우 단순한 개념에서부터 매우 복잡한 개념인 “어린이 보호 법률”까지 다양한 범위를 갖는다(ISO 25964-1 2011, 3). 또한, 용어와 개념 사이에 존재할 수 있는 가능한 등가관계를 고려할 때 이에 대한 구분은 매우 중요하다. 다언어 시소러스를 구축할

때 용어와 개념 구분은 언어 사이에 더 큰 명확성과 등가성을 만든다(Ryan 2014, 13).

기존의 국제규격(ISO 2788)과 현재의 국제규격(ISO 25964)의 기본적인 차이점이라 할 수 있다. 개념의 표현은 용어 등으로 나타낼 수 있으며, 따라서, 용어간의 등가관계(동의어)는 가능하지만, 개념간의 등가관계는 성립하지 않는다.



<그림 1> ISO 25964-1에 의해 정의된 개념과 용어(Ryan 2014, 12에서 인용)

5. 선행 연구

패킷 유형 개발 및 구조화에 대한 연구는 랭가나단의 PMEST 기본 패킷에 기반하여 이를 재활용하거나 확장하는 연구가 많다. 심지영(2014)은 랭가나단의 PMEST 기본 패킷에 기반하여 보도 장르에 적합한 기본 패킷을 누가, 무엇을, 어떻게, 어디서, 언제를 생성하여 보도 장르의 형식적 구조와 내용적 요소를 반영하여 패킷의 구성요소를 추출하였다. 김경민(2015)은 미성년자의 관련 판례정보를 중심으로 법률정보 패킷 체계 개발에 관한 연구를 수행하면서 PMEST를 확장하여 기본 패킷 22개와 하위 패킷 24개, 총 46개의 패킷 체계를 설정하였다.

영역별 패킷 유형의 개발 및 구조화에 대한 연구로 유영준(2008)은 신학 분야의 시소러스를 구축하는 전 단계로서, 패킷을 이용하여 신학 용어를 기본 패킷 7개와 하위 패킷 14개로 분류하였다. 정연경(2013)은 한식 분야의 패킷 구조를 기본 패킷 16개와 하위 패킷 85개를 개발하였다. 인용 순서는 '음식 종류'를 중심으로 재료, 에너지, 공간, 시간 순으로 조합하도록 제안하였다. 연구자는 패킷 분석을 통한 주제 분야별 분류체계 모형 개발의 기초가 될 것으로 기대하고 있다.

최윤경과 정연경(2015)은 기존 주제 접근 도구에서 적용된 다양한 패킷 유형을 분석하여, 우리나라 주제명 표목에 적용 가능한 패킷을 제안하였다. 문헌연구와 사례연구(FAST, PRECIS, Aitchison, Zolon, Bliss, UCD)를 통해 상위 패킷 11개(토픽, 사건, 장소, 시대, 개인명, 단체명, 표제명, 형식, 장르, 언어, 인물)와 하위 패킷 22개를 확립하였다. 기존의 패킷 분석에 대한 연구로 이재열(2000)은 『Uniclass』의 패킷 분석을 통하여 건설 정보 분류체

계의 개념적 기반에 관한 연구를 수행하였다.

패킷의 활용에 대한 연구는 주로 온라인상의 표시에 관한 연구와 일단의 제어어휘에서의 활용 연구가 있다. 유영준(2004)은 기존 분류체계의 구조상의 한계와 특정성과 망라성과 같은 웹 자원의 특성과 관련된 문제를 지적함과 함께 패킷 분석에 의한 웹 자원 조직의 타당성과 그 장단점을 기술하고 패킷 분석을 이용한 구체적인 방법론 세 가지를 제시하였다. 첫째, 패킷 분석을 기반으로 한 색인시스템, 둘째, 열거형 분류체계를 패킷분류체계로 구조를 수정한 대안, 셋째, 국내 검색 엔진의 주제별 디렉토리의 패킷 모형 등의 방법론들을 검토한 결과, 패킷 분석에 의한 제어어휘 구축이 웹 자원을 조직하는 수단으로 활용될 수 있는 가능성을 제시하였다.

시소러스를 통합하는 연구에 활용한 연구로는 이승민(2010)이 물리학 분야에서 널리 사용되고 있는 시소러스인 PACS와 PIRA를 선정하여 이들 시소러스를 개념적으로 통합할 수 있는 패킷 프레임워크를 구축한 연구가 있다. 이 연구에서는 두 시소러스를 연결하기 위해 물리학 분야 전체를 다룰 수 있는 계층구조를 구축하였으며, 이 계층구조에 패킷 분석을 적용하여 각 주제 사이의 관계를 설정할 수 있는 하나의 지식기반을 제공하였다. 또한, 이 개념기반 패킷 프레임워크를 통해 보다 효율적으로 정보를 조직하고 검색할 수 있는 접근방법을 제안하였다.

손대형과 김태수(1998)는 기존의 분석합성형 분류체계(BC2)로부터 개념과 관계를 추출하여 작성한 자모순 시소러스에 패킷 도구를 삽입하여 용어의 개념과 관계를 정확하게 이해할 수 있도록 제시하여, 기존의 분류체계나 시소러스가 가지는 한계를 개선하고자 하였다.

패킷은 또한 온톨로지 등의 상위 또는 코어 온톨로지로 재사용이 가능하다. 이에 대하여 김학래 등(2004)은 『건설정보분류체계』의 최상위 다섯 개의 패킷인 시설물, 공간, 부위, 공종, 자원을 기반으로 건설정보 온톨로지를 온톨로지 구축 도구인 Protege를 통하여 구축하고, 그 활용 방안을 제안하였다. 이 연구에서는 이를 통하여 의미적 상호운용을 지원해줄 수 있다고 밝혔다.

또한, 박희진 외 등(2015)은 패킷 이론을 기반으로 한 온톨로지 구축 방법을 제안하면서, 스마트폰 온톨로지를 설계하는 과정에서 패킷 분류체계를 적용하였다. 이것은 랭가나단과 CRG의 패킷 분류 이론에 기반하여 유형, 사양, 기능, 작업, 상품, 개인, 기업, 장소, 시간 등 9개의 기본 패킷을 설계하였다.

선행 연구를 종합해볼 때, 주로 패킷 유형의 분석 및 개발에 중점을 두고 있으며 이것들을 실제 활용하는 방안은 주로 분류체계나 웹에서의 조직화에 대한 것이 많다. 패킷 분석 기법을 시소러스에 적용하는 연구는 드물다. 한편, 향후 어휘간의 상호운용성이나, 텍사노미나 온톨로지 등에 의한 어휘의 병합이나 통합 등에 대한 연구가 필요할 것이다.

Ⅲ. 건설 분야 패킷 유형 개발

1. 개발 원칙

건설 분야의 기본 패킷을 확립함에 있어서 본 연구에서는 다음과 같은 개발 원칙을 정하였다.

첫째, 특정 영역이나 분야를 한정하지 않은 일반적인 기본 패킷을 조사 및 분석하여 건설 분야의 기본 패킷과 비교 및 분석, 임시 기본 패킷을 생성한다. 이것은 기존의 패킷에서 누락되었거나 기존의 기본 패킷에서 별도의 패킷으로 재설정할 때 망라성을 확보하기 위함이다.

둘째, 기존의 기본 패킷을 분석할 때 하위의 패킷이나 실제 적용된 용어나 개념이 기본 패킷과 상이하거나 세분이 가능한 경우, 분리하여 별도의 패킷으로 한다. 예를 들어, 『건설정보분류체계』에서 “시설물(Facilities)” 내의 도시계획, 단지계획 등의 계획과 개발제한구역이나 수자원보호구역 등의 구역, 산업단지, 유통단지 등의 단지는 별도의 패킷이다.

셋째, 『건설정보분류체계』의 경우 전조합의 분류체계이므로, 분류기호와 캡션을 수집 및 분석할 때, 후조합용의 용어를 상징하고 수집한다. 예를 들어, 부위(Elements) 분류의 “81. 가구”는 부위로서의 가구이고, 공중(Works) 분류의 “57 가구 및 집기”는 공중으로서의 가구이다. 또한, 자재 분류의 “71. 가구”는 자재로서의 가구이다. 그러나 용어 및 개념의 수집에서는 인공물로서의 “가구”를 지칭하는 것으로 수집·분석한다. 이는 시소러스에서는 단일의 의미를 확보해야 하기 때문이다.

넷째, 임시 기본 패킷의 경우는 기존의 기본 패킷을 분석하여 하향식 방식으로 확립하고, 용어 및 개념 분석에 따른 기본 패킷의 확립은 상향식 방식을 채택한다. 예를 들어, 위의 가구라고 하는 개념은 본질 속성에 의해 기본 패킷은 인공물이 된다.

다섯째, 주제관련성만 고려하고 범주관련성은 고려하지 않는다. 특히, 탐색 범위의 축소를 위한 “이용 대상”, “매체 유형”, “형식” 등은 고려하지 않는다.

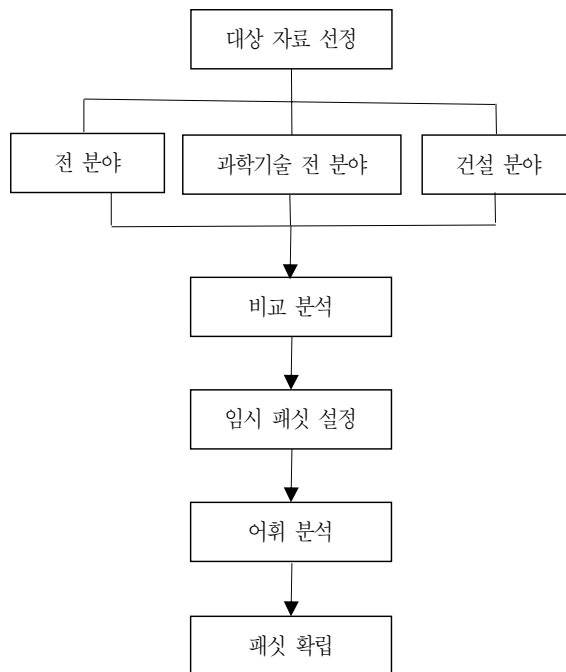
2. 개발 절차

먼저, 문헌조사를 통해 연역적인 방법으로 기본 패킷을 가상(임시)으로 정한다. 특정 영역을 한정하지 않고 설계된 기본 패킷과 과학기술 전반에 적용할 수 있는 기본 패킷, 그리고 건설 영역에 한정된 기존의 기본 패킷을 조사·분석하였다. 예를 들어, 전 영역을 커버하는 Aitchison et al.(2000)의 기본 패킷, 과학기술 분야의 전체를 커버하는 Campbell(1963)의 기본 패킷, 그리고 건설 영역으로 한정된 분류체계에 적용된 『건설정보분류체계』(건설교통부

2006)와 시소러스에 적용된 『건설산업시소러스(CIT)』(Michael 1976) 등이 있다. 본 연구에서는 이들의 기본 패킷들을 우선 비교·분석하여 임시 패킷을 설정하였다. 이 임시 패킷이 설정되고 나면 기존의 『건설정보분류체계』에서 용어 및 개념을 수집·분석하여, 임시 패킷을 보완하고 건설 분야 기본 패킷을 확립하였다.

확립된 패킷을 가지고, 시소러스 구축 표준에서 제시하고 있는 시소러스의 구축 절차에 따라 패킷 분석 기법을 활용한 건설 시소러스의 구축 방안을 제시하였다.

이들에 대한 절차를 그림으로 나타내면 <그림 2>와 같다.



<그림 2> 패킷 유형 개발 절차도

3. 절차별 개발 방안

가. 분석 대상 자료의 선정

기존의 기본 패킷을 조사 및 분석하기 위하여, 특정 영역을 한정하지 않은 전 주제영역, 과학기술 전반, 그리고 건설 영역으로 한정된 기본 패킷으로 나누었다. 이에 따라 전 영역은 Aitchison et al.(2000)의 기본 패킷, 과학기술 분야는 Campbell(1963)의 기본 패킷, 건설 영역은 분류체계인 『건설정보분류체계』 및 시소러스인 『건설산업시소러스(CIT)』를 선정하여 분석하였다.

① Aitchison et al.의 기본 패킷

Aitchison et al.(2000, 70-71)은 오랜 연구 끝에 정착된 기본 패킷을 제시하였다. <표 1>에 일람표를 나타냈다.

<표 1> Aitchison et al.(2000)의 기본 패킷

기본 패킷	하위 구분		예시	
개체/사물/물체	(특성에 따라)	추상적 개체	아이디어, 학문 분야	
		자연 발생 개체	복사, 구름, 강	
		생명체, 유기체	박테리아, 바이러스, 포유동물	
		인공물	제품, 예술품	
		속성: 특성/성질, 상태/상황	온도, 색, 신뢰성, 크기	
		재료/물질, 구성물질	재료, 목재, 가죽, 플라스틱, 직물, 약품	
		부분/구성요소	타이어, 키(방향키), 신체부분, 마루	
	전체/복합체	자동차, 선박, 건물, 학교		
	(기능에 따라)	주체 (행위의 수행자)	개인, 직원, 조직	예술가, 선원, 의사, 협회
		수동체(행위의 수혜자)	설비/장치	공작기계, 컴퓨터, 인공호흡장치, 식기세척기
최종산물			피해자, 수혜자, 오염된 물	
행동/활동	과정 (자동사적 행위)		개발, 질병	
	작동 (타동사적 행위)		매칭, 교수(가르치기), 힐링, 관리	
공간/장소/위치/환경				
시간				

② Campbell의 기본 패킷

Campbell(1963)은 과학기술 분야의 전체를 통해 합리적으로 적용할 수 있는 한 세트의 패킷을 제시하고 있다. <표 2>에 일람표를 나타냈다.

<표 2> Campbell(1963)이 제시한 과학기술 분야의 최상위 패킷

기본 패킷	패킷 정의	예시
대상(thing)	기계, 설비, 장치와 같이 눈으로 볼 수 있는 대상	기계, 설비, 장치
대상의 부분	종종 다수의 다른 오브젝트의 부분을 구성하는 구성요소	타이어, 창문
물질	유의미한 형태를 결한 것으로 대상으로부터 구별된다. 축음기의 레코드는 대상이지만, 이것이 만들어지고 있는 비닐 플라스틱이나 기타의 물질은 대상이 아니다. 가령 물질이 플레이크(flake)나 분말과 같은 형태를 하고 있어도, 통상의 그것은 중요한 것이 아니다.	플라스틱
조작과 작용	인간이 일으키는 조작과, 자연에서 일어나는 것이 확실히 인간에 의해 통제될 것 같은 작용이다. 증류는 이 의미에서의 조작이고, 증발은 그러한 것이 아니	증류, 증발

		다. 다른 공정은 특성 및 현상과 그다지 밀접하게 관련하고 있지 않기 때문에, 한 세트로 하고 있다.	
특성 및 현상		특성은 충식성의 경우와 같이, 종종 하나의 과정을 거치는 능력 혹은 무능력이다. 과정은 최초에 발견되었을 때는 현상이라고 생각되어지는 경향이 있다.	충식성
공간			
시간			
공통세 구분	주제	사회, 법률, 통계와 같이, 특수한 정황 하에서는 그 자체보다도 다른 주제와 관련하여 보다 중요한 주제	사회, 법률, 통계 등의 주제
	문서	서지, 규격이나 학위논문과 같은 문헌의 형식	서지, 규격, 학위논문

③ 『건설정보분류체계』

『건설정보분류체계』(건설교통부 2006)는 시설물분류, 공간분류, 부위분류, 공중분류, 자재분류, 장비분류, 인력분류의 7개 패킷으로 구성되어 있으며, 그 분류체계의 구성은 <표 3>과 같다.

<표 3> 『건설정보분류체계』의 기본 패킷

패킷기호	패킷명	패킷 정의	예시
F	시설물 분류	건설활동의 최종 결과물로서의 시설물과 그 주변 환경요인의 사용목적 및 기능	계획구역 및 토지정비, 운송·교통시설, 석유·화학시설 등
S	공간 분류	공간용도와 물리적 부위요소에 의하여 구성되는 면적 또는 체적 및 개념적 구획을 구분한 분류	구조적 시설 구획공간, 개방형태별공간, 기능적 구획 공간 등
E	부위 분류	물리적인 관점에서 시설물의 한 부분으로서 공간을 둘러싸고, 공간의 기능을 지원하는 시설물의 구성요소	부지준비부위, 토공사부위, 포장 및 도로시설부위, 교량시설부위 등
W	공중 분류	기술적으로 시설물의 한 부위를 구성하는 작업단위로서 제반 자원을 동원하여 고안된 기능을 가지도록 하는 작업 및 작업결과의 분류	공사비, 가설부지, 터널 가설, 측량, 계획, 프리캐스트 콘크리트공사 등
M	자재 분류	건설공중에 투입되는 물리적 자원	화약 및 폭발물, 항공기 및 항공 기체구성품 등
Q	장비 분류	작업을 지원하는 장비	불도우저, 굴삭기, 로우더, 골재생산장비, 준설선 등
L	인력 분류	작업을 직접 수행하는 인력	요리보조, 의료요원, 채봉사, 대기관리기술사, 변전전공 등

④ 『건설산업시소러스(CIT)』

『건설산업시소러스(CIT)』(Michael 1976)는 10개의 패킷으로 구성되어 있으며, 시간, 장소, 속성과 측정, 건설 에이전트, 운영과 과정, 재료, 건조물의 부분, 건조물 등 8개의 핵심 패킷이 분류 배열의 기초를 형성하고 있다. 나머지 2개의 패킷은 일반 패킷으로 하나는 기록물의 형태에 대한 것이고, 또 하나는 건설에 관련된 주변주제(학문 분야 및 기술 등)에 대한 것이다. 10개의 패킷은 <표 4>와 같다.

〈표 4〉 『건설산업시소러스(CIT)』의 기본 패킷

패킷기호	패킷명	패킷 정의	예시
A	기록물의 형태	기록을 수록하는 매체	마이크로폼, 도서, 연감, 카탈로그, 신문, 특허 등
B	주변 주제	블리스 서지 분류법의 학문 분류	철학, 과학, 사회과학, 기술, 정보과학 등
C	시간	일시적인 의미의 용어	년도, 가을, 20세기
D	장소	공간적 또는 위치적인 의미의 용어	유럽, 해외
E	속성과 측정	적합한 측정과 함께 건설에서 주제, 과정, 활동의 재료, 건조물 부분, 건조물에 대한 속성	수량, 모양, 강도, 조명
F	(건설) 에이전트	사람	고용인, 숙련공, 건축가, 목공서, 모형
		프로젝트 및 일반 문헌	도안, 계산서, 계약
		건설활동을 수행하는 과정에서 사용되는 장비	톱, 크레인, 컴퓨터, 측량기
G	운영 및 과정	지적활동을 포함한 건설에서 수행되는 활동 ; 과정은 내적 활동을 반영하는 행동 많은 용어가 운영과 과정의 양자 모두에 속함	설계(designing), 쌓기(laying), 발굴, 광고
H	재료	건설에 사용되는 재료	콘크리트, 시멘트, 유리, 강철, 목재
J	건조물 부분	건조물이나 다른 작업물의 모든 물리적 부분	지붕, 벽, 문, 창문틀
K	건조물	건설업의 최종생산물	공장, 병원, 침실, 부엌, 댐, 교량

나. 비교·분석

Aitchison et al.을 기반으로 Campbell, 『건설정보분류체계』 및 『건설산업시소러스(CIT)』를 대비하여 기본 패킷을 비교하였다. 각각의 패킷별로 포함하는 정의나 범위가 다를 수 있지만 완전일치하지 않더라도 등가로 보았고, 범위를 크게 벗어나는 경우, 분리하거나 신설하였다. 〈표 5〉는 대상 자료에서 제시하고 있는 기본 패킷을 Aitchison et al.의 기본 패킷을 중심으로 비교한 것이다.

〈표 5〉 대상 자료의 기본 패킷 비교

Aitchison et al.	Campbell	『건설정보분류체계』	『CIT』
추상물	주제		주변 주제, 기록의 형태
자연발생물	대상		
생물	대상	인력	주체-사람
인공물	대상	시설물, 장비	주체-장비
속성	특성 및 현상	공중	속성과 측정
재료	물질	자재	재료
부품/구성요소	대상의 부분	부위	건조물의 부분
전체물	대상	시설물	건조물
최종산물	문서	시설물	건조물, 주체-문서
과정	조작과 작용	공중	운영 및 과정
조작			
공간/장소/위치/환경	공간	공간	장소
시간	시간		시간

다. 임시 패킷의 설정

기존의 패킷을 분석하여 임시 카테고리 설정하였다. 방법은 기존의 Aitchison et al.의 기본 패킷(개체/사물/물체, 행위/활동, 공간/장소/환경, 시간 등)을 바탕으로 Compbell의 최상위 8개 패킷(대상, 대상의 부분, 물질, 조작과 작용, 특성 및 현상, 공간, 시간 그리고 공통 세구분인 주제와 문서) 그리고 『건설정보분류체계』의 최상위 7개의 패킷(시설물, 공간, 부위, 공중, 인력, 자재, 장비)과 『건설산업시소러스(CIT)』의 10개의 패킷(시간, 장소, 속성과 측정, 에이전트, 운영 및 과정, 재료, 부위, 건조물 그리고 일반범주의 2개)을 비교·분석하여 전 범위를 포함할 수 있는 패킷을 모두 채용하여 목록화하였다. 상위 패킷 13개, 하위 패킷 14개 등 총27개를 설정하였다. 임시 기본 패킷은 <표 6>과 같다.

<표 6> 임시 기본 패킷

상위 패킷	하위 패킷	범위	비고
주체 및 수동체	주체	사람, 조직	개인, 회사 등
	수동체	행위의 수혜자	피해자, 오염된 물 등
인공물	시설	인력, 설비, 설비의 작동, 토지 등을 부분으로 가지는 전체물	운송시설, 항만시설, 수산시설 등
	구조물	시설의 부분으로서 토지에 정착된 구조물. 건물, 교량, 댐, 터널 등	물환경 위의 구조물을 포함한다. 수상가옥.
	장비	기계나 기구	불도저
부품/구성요소	-	벽, 지붕, 창문, 나사 등 구조물 또는 장비의 부분	
자연물	생물	동물, 식물 등의 생명체/유기체	
	무생물	자연물 중 생물이 아닌 것. 돌, 물, 흙 등	
속성	-	특성/성질, 상태/상황	
공중	-	공사의 종류	토공사, 말뚝공사, 강구조물공사 등
부품/구성요소		구조물 및 장비 등의 부분	지붕, 벽, 엔진 등
최종산물		식품, 찌꺼기, 연구결과	
공통세목	문서	기록의 최종산물	건축법령, 학술논문 등
	기록의 형태	표현을 수록하는 매체	
재료	-	시멘트, 유리, 고무, 목재 등	
행위와 현상	행위	주체가 의도적으로 하는 행위	
	현상	참여자가 없는 자연현상 등.	강우, 연소 등
공간/장소/환경	공간	『건설정보분류체계』의 공간 분류	
	장소	유럽, 경기지방	
	환경	육지환경, 물환경, 대기환경	
시간		일간, 주간, 월간, 연간, 역사적 기간	

라. 어휘 분석

패킷의 보안을 위해 『건설정보분류체계』의 7종의 패킷별 용어를 분석하여 필요한 부분을 임시 패킷에 보완하였다. 어휘의 수집 및 분석 방법은 『건설정보분류체계』에서 7개의 패킷별로 용어를 수집하되 전조합을 고려하여 어휘를 분석하였다.

분류체계의 분류기호(Class Number)는 개념을 나타내지만 캡션(caption)은 개념을 완전하게 표현하지 못한다(ISO 25964-2 2013, 54). 이에 7종의 패킷에서 패킷의 구분원리가 다소 다른 것들을 분리해냈다. 예를 들어, 『건설정보분류체계』의 7개의 패킷에서 별도로 처리한 패킷은 다음과 같다.

첫째, 시설물 분류에서는 “0. 계획 구역 및 토지정비”에서 구역, 계획, 단지를 구분하였으며, 우주선 발사대, 인공위성 등 다수의 구조물과 장비를 분리하여 각각의 패킷에 위치시켰다.

둘째, 공간분류는 동일하다.

셋째, 부위분류는 건축구조, 공법을 분리하였고, 각종 설비를 인공물로 전환하였다.

넷째, 공중분류의 경우, 기본적으로 공사의 종류와 작업을 구분하였고, 작업의 예로는 시험, 측량 등을 고려하였고, 공법은 추상물로 처리하였다.

다섯째, 자재분류에서는, 기본적으로 장비를 분리하여 인공물로 하였다. 예를 들어, 약기는 자재(재료)로 하지 않고 인공물의 장비로 보았다.

여섯째, 장비는 동일하되, 인공물의 하위로 하였다.

일곱째, 인력분류는 주체로 설정하였다.

한편, 건설교통부는 다섯째에서 일곱째의 패킷은 자원(Resource)으로 묶어서 건설공중에 투입되는 자원을 물리적 자원인 자재와 작업을 지원하는 장비, 그리고 작업을 직접 수행하는 인력으로 구분하고 있다(건설교통부 2006).

마. 패킷의 확립

패킷에서 구분의 원리가 다른 것들을 분리하였고, 계획, 공법 등 주로 추상물이 추가되었다. 위의 『건설정보분류체계』의 분석을 토대로 건설 시소러스의 구축 시에 적용될 패킷을 확립하였다.

〈표 7〉 건설 시소러스 구축을 위한 기본 패킷

최상위 패킷	하위 패킷	범위	비고
주체 및 수동체	주체	사람, 조직	각각의 역할로 하위 구분
	수동체	행위의 수혜자	
인공물	시설	구조물, 장비, 토지뿐만 아니라 인력 등을 부분으로 가지는 대상(object)	
	구조물	토지에 부착된 인공물. 건물, 교량, 터널, 댐 등	
	장비	토지에 부착되지 않은 기계나 도구. 불도저, 해머 등	
추상물(개념)	구조	일정한 설계에 따라 여러 가지 재료를 엮어서 만든 시설물의 구조	보완한 패킷임
	공법	건축이나 토목과 관련한 일을 하는 기술이나 방법. 해체공법, 발파공법 등.	보완한 패킷임

	프로젝트	각종 프로젝트 및 계획, 도시계획, 토지정비 등	보완한 패킷임
	사건사고	건설 관련 사건과 사고	보완한 패킷임
재료	-	건축자재, 토목자재 등	자연물은 모두 재료로 처리함
부품/구성요소	-	구조물 및 장비 등의 부분	
공종	-	공사의 종류	
속성	-	특성/성질, 상태/상황	온도, 색, 신뢰성, 크기 등
매체	문서	기록의 최종산물	설계도서, 규격문서, 학위논문 등
	기록의 형태	기록을 수록하는 매체	마이크로폼, 연감, 신문 등
프로세스	행위	주체가 의도적으로 하는 행위	
	현상	참여자가 없는 자연현상 등. 강우, 연소	
공간	-	공간, 장소, 위치, 환경	
시간	-	일, 주간, 월, 연, 계절, 역사적 시간	

4. 패킷의 적용 방안

가. 주체 및 수동체(Agent/Patient)

주체는 통상 의도를 가지고 행위를 할 수 있는 행위주체를 말한다. 사람과 조직으로 나눌 수 있으며, 사람은 직업으로 하위 패킷을 구분한다. 조직은 일반적으로 단체를 뜻한다. 대표적으로 “회사”를 들 수 있으며 건설회사, 건축설계사무소 등이 해당한다. 수동체는 행위의 수혜자로서, 피해자, 오염된 물 등을 가리킨다.

『건설산업시소러스(CIT)』에서 주체는 명시적으로 존재하고 있지만, 의미는 다르게 설정하였다. 『건설산업시소러스(CIT)』의 주체는 “장비”와 “문헌”을 포함하고 있지만, 본 연구에서는 “장비”는 인공물, “문헌”은 매체로 분리하였다.

나. 인공물(Artifact)

인공물은 주체가 행위를 통해 만든 것으로 자연물과는 다른 배타적인 물체를 지칭하고 있다. 『건설정보분류체계』에서 시설물은 건설활동의 최종 결과물로서의 시설물과 그 주변 환경요인의 사용목적, 용도 및 기능에 따른 분류를 말한다(건설교통부 2006). 그러나 본 연구에서는 시설과 구조물을 구분하며, 장비도 인공물에 포함하였다. 한편, 시설은 구조물, 장비뿐만 아니라, 인력, 구조물(건물)이나 장비의 작동, 구조물이 장착된 토지를 부분으로 가지는 인공물이다(장인호 2013, 206).

구조물은 토지에 부착된 인공물을 뜻한다. 건물, 교량, 터널, 댐 등이 이에 해당하며, 물 위에 지어진 가옥도 이에 해당한다. 장비는 토지에 부착되지 않은 기계, 기구 등을 말한다. 불도저, 용접기, 해머 등이 이에 해당한다.

다. 추상물(Abstract)

추상물은 구조, 공법, 프로젝트, 사건사고를 하위로 구분하였다. 구조는 일정한 설계에 따라 여러 가지 재료를 엮어서 만든 시설물의 구조를 말하며, 공법은 건축이나 토목과 관련한 일을 하는 기술이나 방법으로서 해체공법, 발파공법 등이 이에 해당한다. 프로젝트는 도시계획, 토지정비계획 등 각종 계획이나 각종 사업을 말한다. 사건사고는 건설관련 각종 사건과 사고를 말한다.

라. 재료(material)

재료는 콘크리트, 시멘트, 고무, 접착제 등 건설에 필요한 각종 자재를 지칭한다. 『건설정보분류체계』에서는 재료에 조달청의 물품목록체계를 적용하는 것을 원칙으로 하고 있으며, 용역, 서비스, 부품, 기계, 의약품 등 국내 산업 전반의 물품을 관리하는 분류체계이다(정준수 2010, 13). 재료는 일반적으로 건설 분야 분류체계 등에서 자재에 해당하며 시설의 부분으로 하지 않는다.

마. 부품/구성요소(Parts/Component)

부품/구성요소는 구조물 및 장비 등의 부분이나 구성요소를 말한다. 예를 들어, 지붕, 벽, 창문, 나사, 엔진 등을 일컫는다. 구조물이나 장비의 부분은 될 수 있지만, 시설의 부분은 될 수 없다.

바. 공종(Works)

『건설정보분류체계』에서 공종은 기술적으로 시설물의 한 부위를 구성하는 작업단위로서 제반 자원을 동원하여 고안된 기능을 가지도록 하는 작업(및 작업결과)의 분류를 말한다(건설교통부, 2006). 본 연구에서는 공종을 공사와 행위를 구분하였으며 행위의 경우는 프로세스 하위 패킷으로 하였다. 공종의 예시로는 토공사, 방수공사 등 건설과 관련된 공사가 있다.

사. 속성(Attribute)

속성은 특성/성질, 상태/상황으로서 온도, 색, 신뢰성, 크기 등을 말한다.

아. 매체(Media)

매체는 문서와 기록의 형태로 구분하였으며, 문서는 설계도서, 규격문서, 학위논문 등 기록의 최종산물이며, 기록의 형태는 문서를 수록하는 매체이다.

자. 프로세스(Process)

프로세스는 행위와 현상으로 구분하였으며 행위는 주체가 의도를 가지고 행하는 것이며, 현상은 참여자가 없이 일어나는 자연현상을 의미한다. 추상적으로 마음속에 개념으로 존재하는 공사(공중)와는 구분하였다.

차. 공간(Space)

공간은 공간용도와 물리적인 부위요소에 의하여 구성되는 면적 또는 체적 및 개념적 구획을 구분한 분류를 말한다(건설교통부 2006). 본 연구에서는 자연경관, 인공경관과 초자연적 또는 전설적인 장소도 포함하며, 강, 호수 등의 물환경, 섬, 산 등의 육지환경 등도 포함한다.

카. 시간(Time)

시간은 일간, 주간, 월간, 연간 및, 계절, 역사적 기간 등도 포함한다.

IV. 패킷 분석 기반 건설 시소러스 구축 방안

1. 시소러스의 설계 방안

시소러스 설계의 수많은 측면은 구축 전에 고려될 필요가 있다. 이러한 프로젝트를 위해 다음과 같은 방안을 고려해야 한다(Ryan 2014, 26).

첫째, 시소러스의 배포에 있어서 인쇄 또는 전자 형식을 고려해야 한다.

둘째, 시소러스의 표시(display)에 대해 자모순, 계층순뿐만 아니라, 패킷에 의한 분류표시 등이 고려되어야 한다.

셋째, 시소러스가 포함할 구성요소를 파악하여야 한다. 노드 레이블과 용어나 개념간의 관계, 주기사항 등을 고려해야 한다.

이들을 전제로 하여, 본 연구에서 시소러스의 설계 방안으로 고려한 사항은 다음과 같다.

첫째, 시소러스는 용어와 개념을 구분하여 구축하는 것을 전제로 한다. 따라서, 개념에 대한 용어의 표현 및 용어에 대한 동의어를 고려한다.

둘째, 주제관련성만을 대상으로 하고 범주관련성은 고려하지 않는다. 범주관련성은 본 연구에서는 대상으로 하지 않지만, 별도 구성하는 것을 상정한다.

셋째, 패킷 분석 기반으로 설계한다. 본 연구의 주요 요소로서 패킷 분석 기법을 적용하는 시소러스를 상정한다.

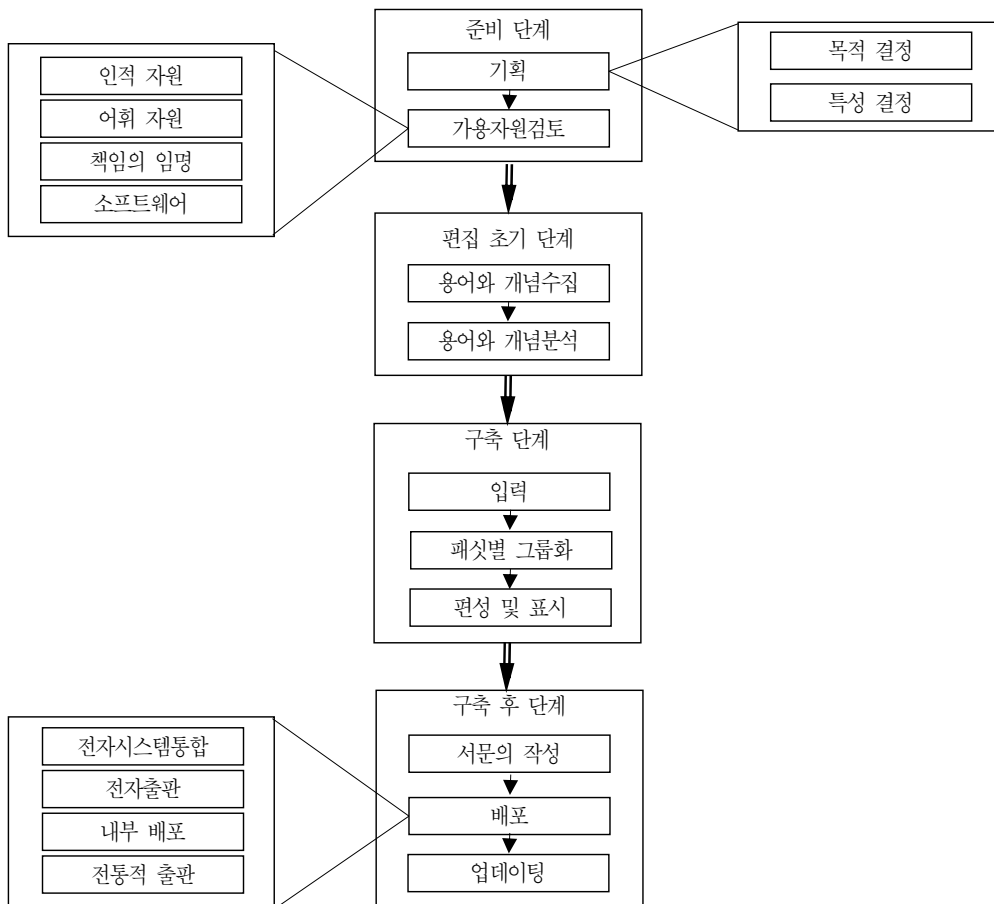
넷째, 시소러스의 구축은 패킷 분석에 의한 하향식으로 출발하지만, 용어와 개념의 수집 및 분석에 따라 상향식을 고려한 하이브리드 구축법을 고려한다.

다섯째, 계층관계는 국제표준(ISO 25964-1 2011)에 따르면서 다중계승을 인정한다. 예를 들어, 피아노는 현악기뿐만 아니라, 건반악기의 하위에도 위치시킨다.

여섯째, 시소러스의 구축 절차는 국제표준(ISO 25964)과 아일랜드 시소러스 구축 지침(Ryan 2014)을 따른다.

2. 구축 절차

본 연구에서는 시소러스 구축 국제표준(ISO 25964-1 2011)과 아일랜드의 시소러스 구축 지침(Ryan 2014)에 따라 <그림 3>과 같은 구축 절차를 따랐다.



<그림 3> 시소러스 구축 절차도

3. 절차별 구축 방안

가. 준비 단계

시소러스의 구축은 시작되기 전에 고려되어야 할 요구의 많은 측면들과 중요한 기획이 요구되는 장기 프로젝트이다. ISO 25964-1은 시소러스에 사용할 수 있는 자원들뿐만 아니라 목표와 기능을 기획하는 것, 구축 및 유지관리의 다양한 측면에 대한 책임의 임명, 소프트웨어를 선택하는 것을 권고하고 있다(Ryan 2014, 26).

① 기획

시소러스의 기획 단계에서는 먼저 시소러스의 목적과 목표, 의도된 이용자, 범위 및 내용에 대해 문서로 작성해야 한다.

② 가용자원의 검토

가용할 수 있는 자원의 검토는 인적자원, 어휘 자원, 소프트웨어, 자금 조달 등이 고려되어야 한다. 인적 자원은 시소러스 편집자, IT전문가, 건설전문인력, 건설관련학자 등으로 구성되는 시소러스 편집팀은 필수이다. 게다가 다국어 시소러스를 고려하는 경우는 해당 언어에 능숙한 인력을 고려해두어야 한다. 이 경우, 편집팀의 일부일 필요는 없지만, 시소러스의 이해를 위해 함께 구성될 수 있다.

어휘 자원은 기존의 어휘 자원은, 택사노미나 주제목록 그리고 기타 키워드 표목과 같은 다른 어휘들이 있다면 사용할 수 있는 경우 참조되어야 한다. 예를 들어, 건설 시소러스에서 『건설정보분류체계』, 『건설산업시소러스(CIT)』 등은 필수적이다. 어휘 자원들은 구축 단계의 용어의 선택 및 기록에서 사용될 수 있다. 기타, 참조도구, 출판된 저작 및 전문가집단과 이용자집단에서 제안하는 용어 등을 고려하여야 한다.

편집자는 초기 단계에서 초판의 편집뿐만 아니라 지속적인 유지보수를 포함하여 모든 단계를 통해 프로젝트를 관리하는 책임을 임명해야 한다.

시소러스의 구축은 소프트웨어 지원 없이 편집하기는 어렵기 때문에 시소러스 관리 소프트웨어를 검토하는 것이 바람직하다.

나. 편집 초기 단계

① 용어와 개념의 수집

기존의 어휘인 『건설산업시소러스(CIT)』, 『건설정보분류체계』, 『건설용어 동의어 및 관련어집』, 기타 시소러스 등에서 건설 카테고리의 기존의 어휘들과 건설업체의 내외부자료, 웹

자원, 전문가집단이나 이용자집단이 제시한 용어나 개념을 가용자원에서 검토된 사항을 반영하여야 한다.

② 용어의 분석

수집된 용어는 그들이 속성을 연구하고 시소러스에 입력할 수 있기 전에 체계적인 순서로 정렬해야 한다. 수집된 용어가 정보원 및 빈도와 함께 기록되면 모든 이형과 동의어가 함께 옮겨지도록 패킷으로 정렬되어야 한다(ISO 25964-1 2011, 91).

다. 구축 단계

① 패킷별 그룹화

패킷에 의해 구조화하는 방법은 상위 클래스 또는 구분들에 따라 그들 스스로의 패킷들을 사용하는 것을 포함한다. 이 방법에서는 용어들은 사물, 재료, 활동 등과 같은 기본 패킷들에 따라 그룹화된다. 이것은 좀 더 상향식 접근법이며, 각각의 패킷 안에서의 모든 용어들은 상위 클래스 또는 구분들과 동일 유형이나 카테고리임을 보증한다. ISO 25964-1은 시간이 지나고 주제가 발전함에 따라 일반적으로 주요 개정이 덜 필요한 이 방식으로 조직화할 것을 언급하고 있다(Ryan 2014, 36).

이후, 범위주기(scope note) 및 개념 간의 관계(계층관계, 연관관계 등) 등의 확립이 필요하다.

② 편성과 표시

범위주기 및 개념 간의 계층 및 연관관계 등이 확립되면, 다음은 편성과 표시이다. 편성과 표시는 소프트웨어의 지원을 받지 않으면 매우 힘든 과정이다. 특히, 시소러스 전용 소프트웨어를 선택하거나 구축할 수 있는 체제를 갖추어야 한다.

라. 구축 이후의 고려 사항

이후의 절차는 전문가 및 이해 관계자의 참여, 시소러스 서문의 작성, 출판 및 배포, 유지와 보수 등을 고려해야 한다.

4. 구축 사례

패킷 분석 기법을 활용한 건설 시소러스의 구축 방안에 따라 시험용 시소러스를 구축하였다. 시험용 시소러스는 약 200여개의 개념을 수집하여 자모순 리스트와 패킷별 리스트를 생성하였다.

가. 자모순 리스트

자모순 리스트는 디스크립터와 비디스크립터를 함께 자모순으로 나열하였다. 비디스크립터는 디스크립터의 안내만을 표시하였고, 디스크립터는 범위주기, 최상위어, 상위어, 하위어, 연관어 등을 나열하였으며, 최상위어를 단서로 패킷별 리스트에서 계층구조를 볼 수 있게 하였다. 자모순 리스트의 예시를 <그림 4>에 나타냈다.

21세기[--世紀] TT 시간 BT 세기	건설관리[建設管理] TT 프로세스 BT 관리행위
가구[家具] TT 인공물 BT 기구	건설기계[建設機械] TT 인공물 BT 기계 NT 로우더 불도저 RT 건설공사
가동교[可動橋] TT 인공물 BT 교량 RT 선박	건설회사[建設會社] TT 주체 및 수동체 BT 회사
가을[--] TT 시간 BT 계절	건자재 → 건축자재
강도[強度] TT 속성 BT 속성	건축공사[建築工事] TT 공중 BT 건설공사
개발행위[開發行爲] TT 프로세스 BT 행위 NT 건축행위	건축구조[建築構造] TT 추상물 BT 구조 NT 절판구조
건물해체공법 → 해체공법	건축도서[建築圖書] TT 매체 BT 문서
건설공사[建設工事] TT 공중 BT 공중 NT 건축공사 조경공사 토목공사 RT 건설기계	건축법규[建築法規] TT 매체 BT 법규
	건축설계사무소[建築設計事務所] TT 주체 및 수동체 BT 회사

<그림 4> 구축 사례(자모순 리스트 일부)

나. 패킷별 리스트

패킷별 리스트는 기본 패킷별로 최상위 패킷을 기준으로 계층 깊이를 알 수 있도록 NT1, NT2 등의 기호 및 들여쓰기(indentation)를 동시에 활용하여 표시하였다. <그림 5>에 그

예시를 나타냈다.

인공물 [NT1] 구조물 [NT2] 교량 [NT3] 가동교 [NT3] 도로교 [NT3] 수로교 [NT3] 운하교 [NT3] 인도교 [NT3] 철도교 [NT3] 혼용교 [NT2] 도로 [NT3] 자동차도로 [NT3] 자전거도로 [NT2] 주택 [NT3] 공동주택 [NT4] 아파트 [NT3] 단독주택 [NT2] 철탑 [NT1] 시설 [NT2] 공장 [NT2] 단지 [NT3] 공업단지 [NT3] 농업단지 [NT3] 복합단지 [NT3] 산업단지 [NT3] 유통단지 [NT3] 주거단지 [NT4] 공동주택단지 [NT5] 아파트단지 [NT4] 단독주택단지 [NT1] 장비 [NT2] 기계 [NT3] 건설기계 [NT4] 로우터 [NT4] 불도저 [NT3] 공작기계 [NT3] 신호장치 [NT3] 조명장치 [NT3] 컴퓨터 [NT3] 탈것 [NT4] 선박 [NT4] 자동차 [NT4] 자전거 [NT4] 항공기 [NT2] 기구 [NT3] 가구	재료 [NT1] 건축자재 UF 전자재 UF 건축재료 [NT1] 목재 [NT1] 시멘트 [NT1] 콘크리트 [NT1] 토목자재 UF 토목재료 주체 및 수동체 [NT1] 수동체 [NT2] 노임 [NT2] 오수 [NT2] 피해자 [NT1] 주체 [NT2] 사람 UF 인간 [NT3] 기사 [NT3] 기술사 [NT2] 조직 [NT3] 회사 [NT4] 건설회사 [NT4] 건축설계사무소 추상물 [NT1] 공법 [NT2] 말파공법 [NT2] 해체공법 UF 건물해체공법 [NT1] 구조 [NT2] 건축구조 [NT3] 절판구조 [NT2] 기본구조 [NT2] 이차구조 [NT3] 바닥 이차구조 [NT3] 방화구조 [NT3] 지붕 이차구조 [NT1] 사건사고 [NT2] 사건 [NT3] 부실공사 [NT2] 사고 [NT3] 공사장 사고 [NT4] 공사장 붕괴사고 [NT3] 붕괴사고 [NT4] 공사장 붕괴사고
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<그림 5> 구축 사례(패싯별 리스트 일부)

V. 결론

본 연구에서는 패킷 분석 기법을 적용한 건설정보 시소러스 구축을 상정하여, 기본 패킷을 확립하고, 해당 기본 패킷의 적용 방안 및 시소러스 구축 방안을 제시하였다. 건설 분야의 패킷은 최상위 패킷 11종과 하위 패킷 13종 등 총 24개의 패킷을 확립하였고, ISO 25964-1과 아일랜드 시소러스 구축 지침의 시소러스 구축 절차에 따른 구축 방안을 제시하였다.

기본 패킷은 주체와 수동체, 인공물, 추상물, 재료, 부품/구성요소, 공중, 속성, 매체, 프로세스, 공간, 시간 등이다.

주체와 수동체는 통상 행위를 할 수 있는 주체와 행위의 수혜자인 수동체의 하위 패킷을 가지며, 주체는 사람과 조직을 말하고, 수동체는 피해자, 오염된 물 등을 말한다.

인공물은 인력, 설비, 설비의 작동, 토지 등을 부분으로 가지는 전체로서의 시설, 토지에 부착되어 있는 설비로서의 구조물, 그리고 토지에 부착되지 않는 장비로 구분하였다. 인공물과 대비되는 자연물은 사람이 만들지 않은 나무, 돌 등을 말하며, 건설 분야의 특수성을 고려하여 재료에 포함하였다.

추상물은 구조, 공법, 프로젝트, 사건사고를 하위 패킷으로 구분하였다. 구조는 일정한 설계에 따라 여러 가지 재료를 엮어서 만든 시설물의 구조를 말하며, 공법은 건축이나 토목과 관련한 일을 하는 기술이나 방법으로서 해체공법, 발파공법 등이 이에 해당한다. 프로젝트는 도시계획, 토지정비계획 등 각종 계획이나 각종 사업을 말한다. 사건사고는 건설관련 각종 사건과 사고를 말한다.

재료는 건설 분야의 분류체계나 시소러스 또는 패킷에서 말하는 통상 토목재료, 건축재료 등의 자재를 말하고, 부품/구성요소는 구조물 및 장비 등의 부분을 말한다. 예를 들어, 지붕, 벽, 창문, 나사, 엔진 등을 일컫는다. 공중은 토공사, 방수공사 등 건설과 관련된 공사의 종류 등을, 속성은 특성/성질, 상태/상황으로서 온도, 색, 신뢰성, 크기 등을 말한다.

매체는 문서와 기록의 형태로 구분하였으며, 문서는 설계도서, 규격문서, 학위논문 등 기록의 최종산물이며, 기록의 형태는 문서를 수록하는 매체이다.

프로세스는 행위와 현상으로 구분되며, 전자는 주체가 의도를 가지고 행하는 행위, 후자는 참여자가 없는 자연현상 등을 말한다.

공간은 장소, 위치, 환경 등을 말하며, 시간은 일간, 주간, 월간, 연간 등과 21세기와 같은 역사적 시간도 포함한다.

패킷의 분석 방안으로는 특정 영역을 한정하지 않은 전 영역을 대상으로 하는 Aitchison et al.(2000)의 기본 패킷, 과학기술 분야의 전 주제를 아우를 수 있는 Campbell(1963)의

기본 패킷, 건설 분야로 한정하고 있는 『건설정보분류체계』에 적용된 기본 패킷과 『건설산업 시소러스(CIT)』에 적용된 기본 패킷을 비교·분석하여 임시 기본 패킷을 수립하고, 『건설정보분류체계』의 어휘를 반영하여 보완하였다.

패킷 분석 기반 건설 분야 시소러스 구축 방안으로는 국제표준에서 제시하고 있는 구축 절차를 바탕으로, 준비 단계, 초기 편집 단계, 구축 단계로 나누어 각각 구축 방안을 제시하고, 구축 이후의 고려해야 할 사항들에 대해 간략하게 설명하였다.

준비 단계에서는 시소러스의 목적이나 특성을 결정하는 기획과 인적자원, 어휘자원, 책임의 임명, 소프트웨어의 선택 등 가용자원을 검토하는 것이 바람직하다. 편집 초기 단계에서는 건설 분야의 기존의 어휘들(분류표, 시소러스 등)을 확보하여, 용어 및 개념을 수집하고 분석한다. 구축 단계에서는 미리 확립된 패킷별로 용어와 개념을 구조화 및 그룹화하고, 소프트웨어를 활용하여 편성하고 표시한다.

구축 이후의 단계에서는 전문가집단에 의한 검토의 실시, 출판 및 배포, 유지와 갱신에 대한 고려가 필요하다는 것을 제시하였다.

구축 방안에 따른 시험용 시소러스는 약 200여개의 개념 및 용어를 수집하여, 자모순 리스트와 패킷별 리스트를 작성하였다.

패킷 분석 기법은 자체 시소러스의 구축에서 뿐만 아니라, 타 제어어휘와의 상호운용성, 타 시소러스와의 병합(merge)이나 통합(integration) 등에 활용할 수 있으며, 특히, 전조합시스템인 분류체계의 『건설정보분류체계』와 후조합시스템인 패킷 기반 시소러스의 상호운용성을 기대할 수 있다. 또한, 향후 온톨로지를 구축하면 온톨로지에 통합됨으로서 활용될 수 있을 것이다. 향후, 이들에 대한 연구와 실제의 구현이 이루어져 활용되어야 할 것이다.

참고문헌

- 김경민. 2015. 『법률정보 패킷 체계 개발에 관한 연구 : 미성년자 관련 판례정보를 중심으로』. 석사 학위논문, 대구가톨릭대학교 대학원 도서관학과.
- 김학래 외. 2004. 통합건설정보분류체계 기반 건설정보 온톨로지 구축. 『한국전자거래학회지』, 9(3): 95-112.
- 박희진 외. 2015. 온톨로지 모델링에서 패킷 분석 활용 연구. 『한국도서관·정보학회지』, 46(2): 257-287.
- 손대형, 김태수. 1998. 패킷분류체계를 이용한 시소러스작성에 관한 연구. 『한국정보관리학회 1998년도 제5회 학술대회 논문집』, 235-238.

- 심지영. 2014. 패킷분석 기법을 적용한 방송자료의 내용 구조화에 관한 연구: 시사보도 뉴스 프로그램을 대상으로. 『정보관리학회지』, 31(3): 313-329.
- 유영준. 2004. 패킷 분석을 이용한 웹 자원의 조직. 『한국비블리아학회지』, 15(1): 23-41.
- 유영준. 2008. 신학 용어 분류를 위한 패킷 설계에 관한 연구. 『한국문헌정보학회지』, 42(3): 259-279.
- 이승민. 2010. 시소러스 통합을 위한 개념기반 패킷 프레임워크 구축. 『한국도서관·정보학회지』, 41(3): 269-290.
- 이재열. 2000. 건설정보 분류체계의 개념적 기반에 관한 연구 : Uniclass의 패킷 분석을 중심으로. 『대한건축학회논문집:계획계』, 145: 3-10.
- 장인호. 2013. “시설”의 온톨로지적 개념 정의에 관한 연구. 『한국비블리아학회지』, 24(3): 199-216.
- 정연경. 2013. 한식 정보 조직을 위한 패킷 구조화에 관한 연구. 『한국문헌정보학회지』, 47(1): 15-37.
- 정준수. 2010. 『건축자재 분류체계의 개선 방안에 관한 연구 : 설계사무소의 작업을 중심으로』. 석사학위논문, 세종대학교 대학원 건축학과.
- 최윤경, 정연경. 2015. 한국 주제명 표목의 패킷 유형 개발에 관한 연구. 『한국문헌정보학회지』, 49(4): 179-201.
- 최윤경. 2015. 『주제명표목의 패킷 구조화 모형 개발에 관한 연구』. 박사학위논문, 이화여자대학교 대학원 문헌정보학과.
- 한국건설기술연구원. 1996. 『건설용어 동의어 및 관련어집』. 서울: 한국건설기술연구원.
- 건설교통부. 2006. 『건설정보분류체계 적용기준』. 건설교통부 공고 제 2006 - 281호.
- Hunter, E. J. 2015. 『분류란 무엇인가 : 지식의 구조화와 검색에 관한 이해』. 박지영 옮김. 서울: 한울아카데미.
- Aitchison, Jean., A. Gilchrist and D. Bawden. 2000. *Thesaurus construction and use: A practical manual*. 4rd ed. Chicago: Fitzroy Dearborn Pubsher.
- Campbell, D. J. 1963. “Making Your Own Indexing System in Science and Techonology.” *ASLIB Proceeding*, 15(1963): 282-302.
- Hunter, E. J. 2015. *Classification Made Simple*. Park, Ziyoung Translator. Seoul: Hanul Academy.
- IFLA Working Group on the Functional Requirements for Subject Authority Records (FRAR). 2010. *Functional Requirements for Subject Authority Data(FRSAD) A Conceptual Model*. International Federation of Library Associations and

- Institutions (IFLA).
- ISO 25964-1. 2011. *Information and documentation— Thesauri and Interoperability with other vocabularies Part 1: Thesauri for information retrieval*. Switzerland: ISO.
- ISO 25964-2. 2013. *Information and documentation— Thesauri and Interoperability with other vocabularies Part: 2 Interoperability with other vocabularies*. Switzerland: ISO.
- ISO 2788. 1986. *Documentation— Guidelines for the establishment and development of monolingual thesauri*. Switzerland: ISO
- Lancaster, F. W. 1972. *Vocabulary Controlled Vocabulary for Information Retrieval*. Washington, D.C.: Information Resource Press.
- Michael, J. Roberts. 1976. *Construction Industry Thesaurus*. 2nd ed., London: Department of the Environment, Property Services Agency.
- NISO. 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. U.S.A: National Information Standards Organization.
- Ryan, Catherine. 2014. *Thesaurus construction guidelines: An introduction to thesauri and guidelines on their construction*. Dublin: Royal Irish Academy and National Library of Ireland.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Chang, Inho. 2013. "A Study on Ontological Conceptual Definition of "Facility". " *The Korean Biblia Society for Library and Information Science*, 24(3): 199-216.
- Choi, Yoon Kyung and Chung, Yeon-Kyoung. 2015. "A Study on Developing Facets for Subject Headings in Korea." *Journal of the Korean Society for Library and Information*, 49(4): 179-201.
- Choi, Yoon Kyung. 2015. *A Study of Developing a Facet Application Model of Subject Headings in Korea*. Ph. D. diss., The Graduate School Ewha Womans University, Korea.
- Chung, Yeon-Kyoung. 2013. "A Study on Structure of a Faceted Classification for Organizing Korean Food Information." *Journal of the Korean Society for Library and Information*, 47(1): 15-37.

- Jung, Jun Soo. 2010. *A Study on the Ways of Improving Classification System of Architectural Material: Based on the Practice of Architectural Firms*. M. A. thesis., Sejong University, Korea.
- Kim, Hak Lae et al. 2004. "Building an Ontology based on the Unified Construction Information Classification System." *The Journal of Society for e-Business Studies*, 9(3): 95-112.
- Kim, Kyung Min. 2005. *A Study on Facet Development of Structure for Legal Information: Focused on Precedents Related to Minor*. M. A. thesis., Catholic University of Daegu, Korea.
- Korea Institute of Civil Engineering and Building Technology. 1996. *The Establishment of Equivalence and Associative Relationship for the retrieval of Construction Terms*. Seoul: Korea Institute of Civil Engineering and Building Technology.
- Lee, Jae-Yeol. 2000. "A Study on the Conceptual Basis of the Construction Information Classification Systems: Focused on the facets of the Uniclass." *Journal of the architectural institute of Korea*. 145: 3-10.
- Lee, Seungmin. 2010. "Construction of the Concept-Based Faceted Framework for Thesaurus Integration." *Journal of Korean Library and Information Science Society*, 41(3): 269-290.
- Ministry of Construction & Transportation. 2006. "Standards for the Application to Construction Information Classification". Public Notification, No. 2006-281.
- Park, Heejin et al. 2015. "A Study on the Use of Facet Analysis for Ontology Modeling." *Journal of Korean Library and Information Science Society*, 46(2): 257-287.
- Shim, Jiyoung. "A Faceted Classification Analysis of TV content: Using News and Current Affairs Programs" *The journal of the Korean society for information management*, 31(3): 313-329.
- Son, Dae-Hyung and Kim, Tae-Soo. 1998. "A Study on Thesaurus Construction Using Facet Classification." *The journal of the Korean society for information management*, 235-238.
- Yoo, Yeong-Jun. 2004. "A Study on Organizing the Web Using Facet Analysis." *The Korean Biblia Society for Library and Information Science*, 15(1): 23-41.
- Yoo, Yeong-Jun. 2008. "Establishing Facet for Classifying Theological Terms." *Journal of the Korean Society for Library and Information*, 42(3): 259-279.

