



# **KIEAE** Journal

Korea Institute of Ecological Architecture and Environment

## 한국생태환경건축학회 논문집의 연구경향 분석 - 최근 10년간 기고논문을 중심으로 -

Research Tends of the Korea Institute of Ecological Architecture and Environment Journal - Focused on Articles Published in the Decade from 2010 to 2019 -

김선혜\*

Sean Hay Kim

\* Corresponding author, Professor, School of Architecture, Seoul National University of Science and Technology, Korea (seanhay.kim@seoultech.ac.kr)

#### ABSTRACT

Purpose: The KIEAE Journal has offered a platform where scholars and practitioners discuss research issues and explore resolutions concerning sustainable environment and ecological architecture for last two decades. Upon publishing the 100<sup>th</sup> volume, the editorial board of the KIEAE Journal initiated analyzing the research trends of the KIEAE Journal to reorganize and modify the existing research area structure and to rebuild research information system to further enclose a wider spectrum of current research issues and case studies concerning sustainable city and architecture. Method: First this study performed desriptive statistical analysis of articles per each research area, and then used a text mining technique, A priori algorithm, to capture causality and latent coherence between keywords. Result: The analysis suggests that the research areas for which only a few manuscripts were published can be restructured. Also it suggests integrating research areas that have similar titles and/or whose actual research contents and methods are not much distinguished from each other. Unfortunately, text mining that intended to capture the keyword causality was not successful, because too many and diverse keywords may degrade learning effectiveness of the A priori algorithm. This study suggests using a group of predefined standard keywords, such that authors can pick up useful keywords out of the keyword pool.

#### © 2020. KIEAE all rights reserved.

#### KEYW ORD

연구경향 전공 분야 키워드 텍스트 마이닝 아프리오리

Research Trend Research Area Keyword Text Mining A Priori

#### ACCEPTANCE INFO

Received May, 18, 2020 Final revision received Jun. 8, 2020 Accepted Jun. 11, 2020

#### 1. 서론

#### 1.1. 연구의 배경 및 목적

한국생태환경건축학회는 2001년 창립되어 그해부터 한국생태환 경건축학회 논문집(KIEAE Journal)을 발행하고 있으며 2020년 5 월 현재까지 총 1501편의 논문을 발행하였다. 또한 KIEAE Journal 은 2006년 한국연구재단 등재후보지로 선정되었으며, 2009년에는 한국연구재단 등재지로 선정되어 매년 격월로 6권, 권당 약 15편의 논문을 게재하고 있다.

KIEAE Journal은 건축학, 건축공학, 건설공학 전반의 생태환경 건축에 관련된 연구와 인문사회학 관점에서의 생태환경건축에 대한 연구를 학계와 산업계에 소개하고, 자연과 인간의 상호관계 및 생태 계를 고려하는 생태환경건축과 지속 가능한 생태환경건축의 보급 방안에 대해[1] 관련 학자와 전문가들이 심도 있는 논의와 해결방안 을 모색 할 수 있는 연구 플랫폼을 제공하고 있다.

KIEAE Journal의 편집위원회는 2019년 12월 통권 제100호 발행 을 계기로 KIEAE Journal에 게재된 논문의 연구경향을 분석하여 발 표하고, KIEAE Journal이 국내외 생태환경건축과 관련된 보다 폭 넓고 현시성 있는 연구 및 사례를 수용할 수 있도록 기존의 연구 분 야 체계와 연구 정보 체계를 정비 · 개선하고자 본 연구를 진행하였 다.

#### 1.2. 연구의 방법 및 범위

본 연구는 KIEAE Journal 게재 논문의 세부 전공 분야 분석과 키 워드 연관성 분석의 두 부분으로 나누어져 있다. 현재의 연구 분야 체계가 정립된 2014년 이후 게재된 논문을 세부 전공 분야 분석 대 상으로, KIEAE Journal이 한국연구재단 등재지로 선정된 이후인 2010년부터 2019년까지 10년간 게재된 논문을 키워드 연관성 분석 대상으로 설정하였다. 이에 따라 약 900여 편 게재 논문의 제목. 키 워드, 초록, 세부 전공 분야 데이터를 수집하였다. 세부 전공 분야에 는 기술적인 통계 분석을 수행하였지만, 키워드 연관성 분석은 기계 학습 텍스트 마이닝을[2] 활용하여 키워드 간의 인과관계와 숨어있 는 규칙을 분석하여 잠재적인 연구 경향을 추론하고자 하였다.

#### 2. 연구 분야 분석

#### 2.1. KIEAE Journal의 연구 분야

Table 1.에 나열되어 있듯이 투고자가 선택할 수 있는 KIEAE Journal의 세부 전공 분야에는 크게 건축학, 건축공학, 실내건축학, 도시공학, 사회문화공학 분야가 있으며, 건물 에너지, 건축 환경 및

Table 1. Research areas by KIEAE journal as of 2014

	· · · · · · · · · · · · · · · · · · ·
Research areas	Sub-areas
Architecture	Architectural planning, Architectural site planning, Environmental design
Interior architecture	Interior architecture
Urban landscape planning	Urban planning, Landscape planning, Urban site planning
Architectural engineering	Material, Construction, Structure
Built environment	Built environment planning, Building environmental control, Building energy, Thermal environment, Visual environment, Acoustic environment, Indoor air quality
Cultural and social science	Cultural science, Social science

건축 환경제어에 특화된 연구는 동일한 이름의 해당 세부 연구 분야로 투고 가능하다[3]. 이는 '생태환경건축'의 연구범위가 환경계획이나 건축환경설비 같은 특정한 학문영역보다 광범위하기 때문이며, 공학뿐만 아니라 미학, 경제학, 사회문화, 법률 등 인문사회학의관점에서도 자연/생태계/인간간의 상호관계 및 지속가능한 구조물이 연구 주제로 다루어지기 때문이다. 따라서 논문을 투고할 때 세부전공 분야의 다중선택이 가능하다. 만약 투고 논문과 적합한 연구 분야가 없다고 판단될 때, 투고자가 기타 분야로 지정하고 직접 연구분야를 기입할 수 있다.

#### 2.2. 연도별 세부 전공 분야 비율 추이

Fig. 1.에서 Fig. 6.은 2014년에서 2019년에 게재된 논문 중 해당 년도에 세부 전공 분야별 게재된 논문의 비율을 나타낸다. 이를 자세히 살펴보면:

- 1) 건축계획과 건물에너지 분야로 투고한 논문의 비율이 1, 2위를 차지하고 있으며, 최근 들어 건축계획 분야로 투고한 논문의 비율이 점점 늘어나고 있는 추세이다. 그 이유는 건물에너지 분야의 국내 논 문집이 늘어나고 있어 투고 논문들이 분산되기 때문이라 사료된다.
- 2) 뒤를 이어 건축 환경계획, 건물 열환경, 도시계획, 건축 환경설계, 건물 환경제어 및 건물 공기환경 분야로 투고한 논문의 비율이 높은 것으로 나타났다.
- 3) 실내건축, 건축공학, 건물 빛환경 및 음환경 분야로 매년 수편 의 논문이 투고되었으나, 대상기간 중 실내건축 및 사회문화공학 분 야로 투고된 논문은 매해 한편 정도에 그쳤다.
- 4) 기타 분야로 지정된 논문도 매년 한편정도 게재되었다. 그러나 연구 내용으로 볼 때 Table 1.의 세부 전공 분야에 속한다고 볼 수 있 었기 때문에, 세부 전공 분야를 변경 지정하여 분석에 포함하였다.

#### 2.3. 세부 전공 분야별 연도에 따른 추이

Fig. 7.에서 Fig. 10.은 2014년에서 2019년까지 게재된 논문 중 세부 전공 분야별 논문이 해당 연도의 전체 논문 중 차지하는 비율을 나타낸다. 2.2절에서의 분석하였던 연구경향을 다시 확인할 수 있다.



Fig. 1. Share by research areas in 2014(total 41 articles)



Fig. 2. Share by research areas in 2015(total 58 articles)

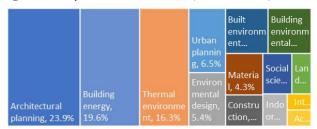


Fig. 3. Share by research areas in 2016(total 92 articles)



Fig. 4. Share by research areas in 2017(total 96 articles)

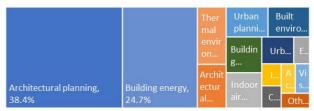


Fig. 5. Share by research areas in 2018(total 73 articles)



Fig. 6. Share by research areas in 2019(total 91 articles)

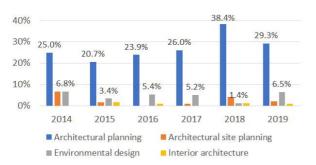


Fig. 7. Share by research areas(Architecture and Interior architecture) from 2014 to 2019



Fig. 8. Share by research areas(Urban and landscape planning, and Architectural engineering) from 2014 to 2019

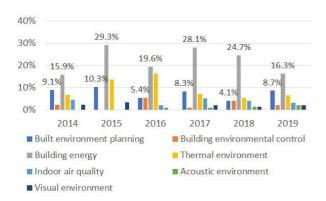


Fig. 9. Share by research areas(Built environment) from 2014 to 2019

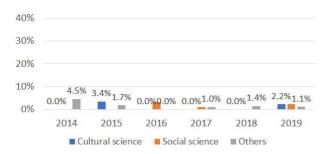


Fig. 10. Share by research areas(Cultural and social science and others) from 2014 to 2019

투고자가 세부 전공 분야를 중복 선택할 수 있기 때문에 적용된 세부 전공 분야수를 분석하였을 때, 하나의 세부 전공 분야에만 해당하는 논문은 전체의 87.3%였다(Fig. 11.).

전체 논문 중 약 13%에 해당하는 48편의 논문이 두 가지 이상의

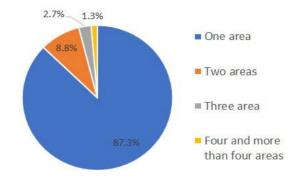


Fig. 11. Share by the number of(duplicated) research areas

세부 전공 분야에 해당되었다. 이 논문들을 분석하였을 때, 건축계획, 환경설계, 환경계획, 건물에너지, 열환경 세부 전공 분야가 가장 빈번하게 중복으로 선택되었다. 이는 2.2절에서 나타났듯이 비교적 많은 논문들이 투고되었던 세부 전공 분야와 일치한다. 중복 전공 분야 논문들을 좀 더 세부적으로 살펴보면:

1) 환경계획과 건축계획 분야가 동시에 선택되는 경우가 빈번히 있었다. 환경설계와 환경계획이 중복으로 선택되는 경우도 빈번히 있었다.

2) 건물 에너지나 열·빛· 공기·음 건축환경 중의 한 분야가 분석 대상 성능일 경우 건축계획이나 환경계획 분야가 중복으로 선택되는 경우가 빈번히 있었다.

3) 건물 에너지와 열환경 분야가 중복으로 선택되는 경우가 가끔 있었다. 두 가지 이상의 환경 성능이 분석될 경우 해당 전공 분야가 중복으로 선택되었다.

4) 환경 성능 분석 대상이 도시나 조경일 경우 해당 분야가 중복 선택되었고, 건축물의 재료나 구조, 시공방법이 환경 성능과 연관된 경우에도 해당 분야가 중복 선택되었다.

5) 건축 단지 계획과 도시 단지 계획이 유사한 전공 분야로 보이지 만, 연구 대상이 건축물이면 건축 단지 계획이 중복 선택되었고 연구 대상이 도시이면 도시 단지 계획이 중복 선택되었다.

6) 사회문화공학 분야로 투고된 논문은 건축계획이나 도시계획 분야와 중복 선택되는 경우도 있었다.

## 3. 키워드 연관성 분석

논문의 키워드는 논문의 제목과 함께 논문이 전달하고자 하는 의도를 가장 축약하여 전달한다. 논문의 제목에서 명사형 키워드를 추출하는 경우도 많지만 타 연구와의 차별화된 특성이나 특정 연구방법론을 키워드로 지정하는 경우도 많기 때문에, 키워드를 분석하면해당 논문의 연구 방향을 쉽게 파악할 수 있을 뿐더러해당 연구 분야의 현안이나 연구자들의 관심도도 쉽게 유추할 수 있다.

특히 학술논문은 투고 당시 적어도 3-5개의 키워드를 제출하도록 되어 있기 때문에 개별 키워드보다 키워드 집합 (Keyword set)이 전달하고자 하는 정보량이 더 다양하고 유의미한 정보를 제공한다. 예를 들어 특정 연구 분야에서 특정 키워드 집합이 자주 발견된다면, 연구 방법론의 경향을 기술적으로 파악하는데 수월하다. 즉, 후속연구자가 관련문헌을 조사하거나 연구 방향을 설정할 때 어떤 키워

드로 검색을 해야 하는지. 어떤 키워드는 피해야 하는지에 대한 가이 드라인으로 키워드간의 연관규칙이 사용될 수 있다.

#### 3.1 키워드 추출 방법

2010년부터 2019년까지 10년간 게재된 논문의 키워드를 추출하 였다. 한 논문 당 적어도 3개 이상의 키워드가 지정되어 있고, 많은 경우 6개의 키워드가 지정된 경우도 있었다. 대부분 논문에서 키워 드 집합이 논문의 의도와 목적을 잘 표현하고 충분한 대표성을 지니 고 있었으나, 일부 논문의 경우 관련이 적거나 과도하게 일반적인 키 워드가 포함된 사례도 있었다. 이러한 경우 관련 있는 키워드를 제목 과 초록에서 추출하여 재구성하였다. 또한 같은 의미를 지닌 키워드 가 여러 형식으로 표기되는 경우가 많아 (예, 에너지 고효율, 에너지 절약, 에너지 고성능, 저에너지) 키워드의 의미를 훼손하지 않는 범 위 내에서 키워드를 수정하였다.

#### 3.2. 텍스트 마이닝

자연어나 책의 문장같은 비정형 텍스트 데이터에서 통계적으로 의미 있는 개념이나 특성을 찾아내고 이들의 패턴이나 경향을 찾아 내는 과정 또는 기술을 텍스트 마이닝이라 한다.

텍스트 마이닝 기법은 크게 단어 빈도 분석, 군집 분석, 감정분석, 연관성 분석으로 나누어지는데, 먼저 전체 키워드 데이터의 흐름을 이해하기 위해 키워드 빈도 분석을 실시하고, 두 개 이상의 키워드가 주어졌을 때 키워드 간 서로 얼마나 연관되어 있는지 분석하는 연관 성 분석을 진행하였다. 순차 연관성이 높은 규칙들이 많이 발견되면 그러한 규칙의 집합은 결국 전체 데이터의 대표적인 패턴이 되어 경 향을 파악하기 쉬울 뿐만 아니라 예측도 용이하다.

#### 3.3. A priori 알고리즘

'A priori'는 경험 없이, 즉 선험적으로 알 수 있는 진리를 의미하 며. 논리법칙이나 수학의 정리가 대표적인 A priori 논리이다. 기계 학습에서 A priori 알고리즘은 대표적인 비지도 학습 알고리즘으로 연관규칙 분석 알고리즘으로 불리우며, 그 이름이 의미하듯이 조건 절(Antecedent)과 결과절(Consequent)로 구성된 규칙을 제시한다.

알고리즘이 제시한 어떤 규칙이 다른 규칙에 비해 얼마나 유용한 지를 판단해 주는 지표로 지지도(Support), 신뢰도(Confidence) 와 향상도(Lift)가 있다.

조건절: A → 결과절: B라는 규칙이 있을 때, 지지도는 A 이벤트 가 일어날 확률을 의미하고(Eq. 1), 신뢰도는 A 이벤트가 일어났을 때 B 이벤트가 일어날 조건부확률(Eq. 2), 지지도는 B 이벤트가 단 독으로 발생할 확률 대비 A 이벤트가 일어나고 B 이벤트가 일어날 확률을 의미한다(Eq. 3).

$$Support(A) = P(A)$$
 (Eq. 1)

Confidence(A 
$$\rightarrow$$
 B) =  $\frac{P(A \cap B)}{P(A)}$  (Eq. 2)

$$Lift(A \to B) = \frac{Confidence(A \to B)}{Support(B)} = \frac{P(A \cap B)}{P(A) \cdot P(B)}$$
 (Eq. 3)

지지도는 전체 데이터에서 조건절 A가 얼마나 발생하는지를 나 타내고, 신뢰도는 조건절 이벤트 A와 결과절 이벤트 B의 연관성 정 도를 나타낸다. 따라서 지지도가 어느 정도 높지 않으면 유용한 규칙 이 아니며, 신뢰도가 높을수록 의미 있는 연관규칙이다. 향상도가 1 이라면(즉, 이벤트 B가 일어날 확률이 이벤트 A가 일어난 후 이벤트 B가 일어날 확률과 같다면) 조건절 이벤트 A와 결과절 이벤트 B는 서로 독립사건임을 의미한다. 즉 규칙 사이에 유의미한 연관성이 없 다는 것이다. 그러나 향상도가 2라면, 이벤트 A가 일어나고 이벤트 B가 일어날 확률이 이벤트 B가(이벤트 A와 상관없이) 단독으로 발 생할 확률보다 2배 높다는 의미이므로 이벤트 A와 이벤트 B의 연관 성이(서로 독립사건일 때 보다 상대적으로-정확히는 2배 정도) 크 다고 볼 수 있다. 따라서 A priori 알고리즘을 실행하면 향상도가 높 은 규칙, 즉 강한 연관성이 있는 규칙부터 먼저 제시한다.

#### 3.4. 텍스트 마이닝 프로세스

3.1절에서 추출한 키워드 집합은 원시 데이터(Raw data)이기 때 문에, 이를 A priori 알고리즘의 학습 데이터로 사용하기 위해서 데 이터 분절(Data tokenization)<sup>1)</sup>과 데이터 정제(Data cleaning)<sup>2)</sup> 등 의 전처리를 먼저 수행하였다. 그 후 개별 키워드로 작성된 데이터 사전(Data dictionary)을 작성하고 키워드별 출현 빈도를 계산하였 다. 전처리가 끝난 키워드 집합 중 지지도가 0.0005 (0.05%) 이상인 키워드 조합만을 A priori 알고리즘의 학습 데이터로 입력하였다.

#### 3.5. 키워드 빈도 및 연관성 분석 결과

3.1절에서 작성한 데이터 사전에서 가장 출현 빈도가 높은 50 키 워드를 추출하여 워드 클라우드로 가시화하였다(Fig. 12.). 키워드 클라우드에서 키워드의 크기는 출현 빈도에 비례하므로, 최근 10

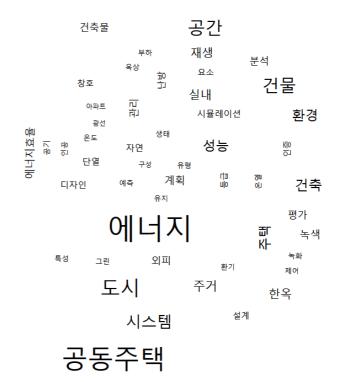


Fig. 12. Keyword cloud

년간 주로 {에너지(76회), 공동주택(71회), 도시(49회), 공간(42회), 건물(41회), 시스템(39회), 환경(34회), 성능(33회), ...}에 관련된 논 문들이 KIEAE Journal에 게재되었다고 볼 수 있다.

A priori 알고리즘으로 키워드간의 연관규칙을 분석한 결과 '비교 적' 빈번히 나타나고 순차 연관성이 있다고 볼 수 있는 키워드의 조 합이 Table 2.에 정리되었다. 여기서 →는 순차 연관성을, ↔은 상호 연관성을 의미한다. 예를 들어 {가로시설물, 친환경가로시설}이 어 떤 논문의 키워드라면 {대체에너지, 신재생에너지}도 그 논문의 키 워드가 될 가능성이 높으며, 그 반대의 경우도 가능성이 높다는 의미 이다. 또한 {방열판}이 어떤 논문의 키워드라면 {LED}도 그 논문의 키워드가 가능성이 높다는 의미이다.

그러나 Table 2.의 연관규칙은 통계적 의미가 없는 것으로 분석되 었는데, 그 이유는 각 연관규칙의 지지도가 0.001 (0.1%) 정도 밖에 되지 않기 때문이다. 그럼에도 불구하고 Table 2.에 나타난 키워드 들이 빈번하게 나타났고 연관성이 크다고 알고리즘이 판단한 이유 는, 동일한 저자가 동일한 연구범위의 유사 논문을 수편 게재했기 때 문에, 1-2편 밖에 특정 키워드 조합이 출현하는 다른 논문에 비해 '비교적' 출현 빈도가 높았기 때문이다.

결론적으로 텍스트 마이닝을 활용하여 키워드 간의 순차 연관성 이나 상호 연관성을 분석하고자 하는 목적은 달성하지 못하였다. 그 이유는 A priori 알고리즘은 키워드가 나타나는 확률을 계산하기 위 해 희소행렬(Sparse matrix)을 작성한다. 키워드가 많을 경우 희소 행렬이 아주 방대해 지고 희소행렬의 셀이 대부분 0이 되어 특정한 키워드 조합이 나타날 확률이 아주 적어진다.3) 따라서 다른 키워드 조합보다 2-3번 정도 밖에 더 나타나지 않았더라도 Table 2.처럼 '비교적' 출현 빈도가 높았던 조합이 상위권에 포진할 가능성이 커 지게 되지만, 결국 '절대적' 출현 빈도가 높지 않기 때문에 해당 키워 드에서 파생된 연관규칙이 통계적으로 의미 있을 가능성이 희박 하다.

따라서 텍스트 마이닝으로 의미 있는 연관규칙을 발견(Mining) 하기 위해서는 1) 한정된 범위의 표준 키워드를 사용하고, 2) 분석대 상이 어느 정도는 동질한 성격을 가지고 있어 연구집단 내에서 사용 하는 어휘가 너무 다르거나 다양하지 않아야 한다.

Table 2. Top ten rules by the a priori algorithm

Rank	Rules
1	{가로시설물, 친환경가로시설} ↔ {대체에너지, 신재생에너지}
2	{방열판} → {LED}
3	{텍토닉} ↔ {산티아고칼라트라바}
4	{철골프리캐스트콘크리트합성보} ↔ {등분포하중, 대칭가 력, 파괴양상, 시공하중}
5	{병실} ↔ {의료법}
6	{노후초등학교, 리모델링} → {에너지절감요소, 에너지시뮬 레이션}
7	{Emergy} → {건물}
8	{전시공간} → {박물관}
9	{베네시안블라인드, 건물에너지성능} ↔ {불쾌현휘, 슬랫각도, 조명제어}
10	{르자연} → {경관}

Table 3. Suggested classification of research areas

Research areas	Sub-areas
Planning and design	Architectural planning, Interior architecture planning, Site planning, Urban planning
	Landscaping, Landscape design, Public design
Architectural engineering	Material, Construction, Structure
Built environment	Built environment planning, Building environmental control, Building energy, Indoor environment (Thermal environment, Visual environment, Acoustic environment, Indoor air quality)

## 4. 결론

최근 10년간 KIEAE Journal에 게재된 논문의 연구경향을 분석하 여 발표하고. 기존의 연구 분야 체계와 연구 정보 체계를 정비 · 개선 하고자 본 연구를 진행하였다. 분석결과 다음과 같은 결론을 얻을 수 있었다.

1) 게재된 논문의 세부 전공 분야를 분석하였을 때, 비교적 게재 가 적었던 세부 전공 분야는 축소하고 세부 전공 분야명이 유사하거 나 실질적인 연구대상이 유사한 세부 전공 분야는 통합을 고려해 볼 수 있다.

이에 따라 KIEAE Journal의 편집위원회는 Table 3.과 같이 세부 전 공 분야를 조정하였다. 건축과 도시 계획 및 디자인을 한 전공 영역 으로 묶고, 현재 많은 연구들이 발표되고 있는 경관 디자인과 공공 디자인 세부 전공을 신규로 추가하였다. 건축환경 전공 분야에서 열 환경이나 공기환경에 대한 게재 논문 비율이 높기는 하지만, 빛ㆍ음 환경과 동일한 연구영역 레벨이기 때문에 건축환경(Indoor Environment) 세부 전공 분야로 통합하였다.

2) 기계학습의 텍스트 마이닝 기법을 활용하여 키워드 연관성 분 석 후 유의미한 결과를 도출하기 위해서는, 투고자가 주어진 키워드 그룹에서 원하는 키워드를 선택하도록 하고 논문의 특수성을 반영 하는 키워드는 기존대로 작성하게 하는 방식을 도입할 것을 권고 한 다. 이 때 키워드 그룹은 원시 키워드 데이터를 군집 분석(Clustering) 하여 대표 키워드를 추출할 것을 추천한다. 또한 텍스트 마이닝의 범 위를 전 전공 분야가 아니라 특정한 전공 분야로 한정하여 진행하고. 각 전공 분야별로 도출된 일정 순위 이상의 연관규칙들을 통합하여 각 연관규칙의 지지도 별로 전체 순위를 매기는 방식이 더 유용할 것 으로 사료된다.

### Acknowledgement

The author has a special thank to Gayoung Kang and Hojin Sung for collecting and assorting data.

This work was supported by the National Research

Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No. 2020R1A2C1012952).

### Reference

- [1] Korea Institute of Ecological Architecture and Environment Journal, Introduction, http://www.kieae.org/11, 2020.05.14.
- [2] Text Mining with R: A Tidy Approach. Julia Silge, David Robinson. O'Reilly Media, Sebastopol, 2017.
- [3] Korea Institute of Ecological Architecture and Environment Journal, Manuscript submission form, http://www.kieae.org/논문투고양식, 2020. 05.14.
- 1) 키워드 내에서 명사와 명사의 분리, 명사와 조사의 분리 등.
- 2) 불필요한 데이터 삭제, 특수문자 및 숫자 삭제, 불용어 제거, 중복단어 제거, 띄 어쓰기 표준화 등
- 3) 지지도가 0.0005 (0.05%) 이상인 키워드 조합만이 A priori 알고리즘의 학습 데이터로 입력된 이유이기도 하다.