https://doi.prg/10.7743/kisee.2019.18.2.001

빅데이터를 활용한 폐교시설의 지표 개발에 관한 연구 -텍스트마이닝 기법을 중심으로-

A Study on the Development of the Use Index of Closed School Facilities Using Big Data -Focused on Text-Mining Techniques-

김 재 영* 이 종 국**
Kim, Jae-Young Lee, Jong-Kuk

Abstract

The purpose of this study is to make objective decisions in the use of closed schools through the development of utilization indicators for the efficient use of closed schools, which is expected to increase continuously. The research phase was largely carried out by drawing preliminary indicators for use in closed schools, drawing final indicators using big data, and quantifying indicators, and finally objectifying them through quantification. The institution intends to apply and verify the facility based on future indicators. This study has implications for the application of big data analysis methods that have not been attempted in planning and research for the use of closed school facilities to date.

키워드 : 빅데이터, 폐교시설, 활용지표, 텍스트마이닝, 정량화
Keywords : Big data, Closed School, Utilization Indicator, Text Mining, Quantification

I. 서론

I-1. 연구의 배경 및 목적

최근 학령인구의 급격한 감소로 지금까지 농어촌 지역에서 주로 발생했던 폐교시설이 수도권을 비롯한 대도시지역에서도 발생하고 있는 현실이다. 이에 정부에서는 폐교 재산의 활용 촉진을 위한 특별법을 마련하는 등 그 대책을 모색하고 있으나 최근 폐교현황자료 기준 전국 3,752개의 폐교 중 420개(11.2%)의 폐교시설이 미활용되어 방치되고 있으며,[1] 임대기간의 만료와 폐교가 예정된 학교로 인해 그 수가 지속적으로 증가할 전망이다.

폐교의 활용은 시설이 위치한 지역적 특성과 지역주민의 의견 등 다방면으로 충분한 계획을 수립하고 검토하여 합리적인 활용방안을 모색해야 하지만 미활용 폐교의 수를 줄이기에 급급하고, 활용을 위한 뚜렷한 기준이 없기 때문에 활용 성공 사례보다는 실패 사례가 더 많은 현실이다. 또한 폐교는 앞으로도 지속적으로 증가할 것으로 전망되므로 실질적인 활용을 위한 대안 마련이 필요한 시점이다.

폐교와 관련된 연구는 꾸준히 진행되고 있으나 접근방법에 있어서는 사례조사 및 설문조사를 중심으로 한 활용방안의 제안으로 폐교활용 시 참고자료로써 기여하지만 연구결과의 실질적인 활용은 기대할 수 없다는 한계가 있다. 또한 설문조사의 경우 표본 및 설문대상의 수, 설문기간 등에 대한 제한이 있고, 신뢰성과 정확성 확보에 어려움이 따른다. 이러한 문제점으로 최근 많은 기업과 정부기관에서는 설문조사와 빅데이터 분석을 병행하거나 정책수립과 민원대응 과정에서 빅데이터의 수집과 분석을 적극적으로 도입하는 등 다양한 정보 수집을 통한 효과적인 대안 마련이 꾸준히 이루어지고 있다. 폐교시설의 활용 계획수립 단계에서도 빅데이터를 활

* 계명대학교 일반대학원 건축학과
** 계명대학교 건축토목공학부 건축학전공

1) 지방교육재정알리미, http:/www.eduinfo.go.kr(2019년 3월 20일 검색).

용한 효과적인 분석 방법을 도입한다면, 기존의 활용방법 보다 다양하고 실질적인 활용대안의 제시가 가능할 것으로 판단된다.

따라서 본 연구는 폐교시설 활용을 위한 객관적이고 합리적인 지표를 개발하는데 최종 목적을 두고 있다. 이를 위해 먼저 관련 연구와 지표를 선정하여 상호비교 함으로써 예비 지표를 도출하고, 이를 빅데이터를 활용하여 검증하여 1차, 2차 지표 도출 과정을 거쳐 최종 지표를 제안하고자 한다. 또한 ANP분석을 통해 활용지표를 정량화함으로써 객관성을 확보하였다.

I-2. 연구의 범위 및 방법

연구 단계는 크게 예비지표 도출 단계, 빅데이터 분석을 통한 최종지표 도출 단계, 지표 정량화 단계로 구분된다. 빅데이터 분석은 도출된 예비지표의 타당성 검증을 위해 폐교 활용과 관련된 보다 방대한 자료의 수집과 분석을 통해 지표의 타당성을 검증하여 객관화된 폐교시설의 활용지표를 제안하는 것이다. 그리고 ANP 분석을 통한 지표 정량화 단계에서는 앞서 제안된 폐교시설의 활용지표를 기준으로 전문가 설문조사를 통해 지표 간 상관성을 검토하고, 우선순위 및 가중치를 도출함으로써 이에 따른 지표 항목을 배점화하였다. 세부내용은 다음과 같다.

첫째, 선행연구의 검토를 위해 선행연구와 관련 연구보고서 등을 근거로 하여 폐교시설의 리모델링과 활용 요소를 중심으로 객관성이 확보된 선행연구를 선정하였다. 국내외 공공디자인 지표는 국내 디자인 서울 가이드라인, 서울시 공공디자인 평가시스템, 행정중심복합도시 좋은 건축물 지정제도를 선정하였고, 국외 영국 DQI(Design Quality Indicators), 영국 건축위원회 디자인리뷰(Design Review)를 선정하였다.

둘째, 빅데이터의 다양한 활용 기법 중 단어간 연계성 파악과 분류를 위한 텍스트 마이닝(Text mining)기법을 적용하였다. 자료수집을 위해 신문기사 데이터베이스 시스템인 BIGKinds를 활용하였으며 폐교재산의 활용촉진을 위한 특별법이 개정·시행된 2014년 3월 11일을 기준으로 2019년 3월까지 폐교시설 활용에 관하여 보도된 중앙지와 지역종합지를 대상으로 2,384개의 기사자료를 수집하였다. 텍스트 마이닝 기법 적용을 위한 분석도구는 오픈소스로써 다양한 빅데이터 분석 관련 연구에 사용되고 있는 'R'프로그램을 주요 분석도구로 하여 데이터 전처리 및 분석에 따른 패키지를 활용하였다.

셋째, ANP(Analytic Network Process, 네트워크 분석적 의사결정법)분석을 통한 우선순위의 도출과 정량화를 위해 상관관계 분석을 바탕으로 네트워크 분석적 의사결정 시뮬레이션을 구현할 수 있는 SuperDecisions 2.8 프로그램을 사용하였다. 지표간 상관관계 검토를 위해 전문가 5인(건축 3인, 디자인2인)을 대상으로 FGI 설문조사를 실시하였다. 중요도 분석은 관련 전문가(건축, 도시, 관련 공무원) 100명을 대상으로 설문조사를 실시하였으며 표본 수는 100부를 배포하여 48부를 회수하였다. 설문지의

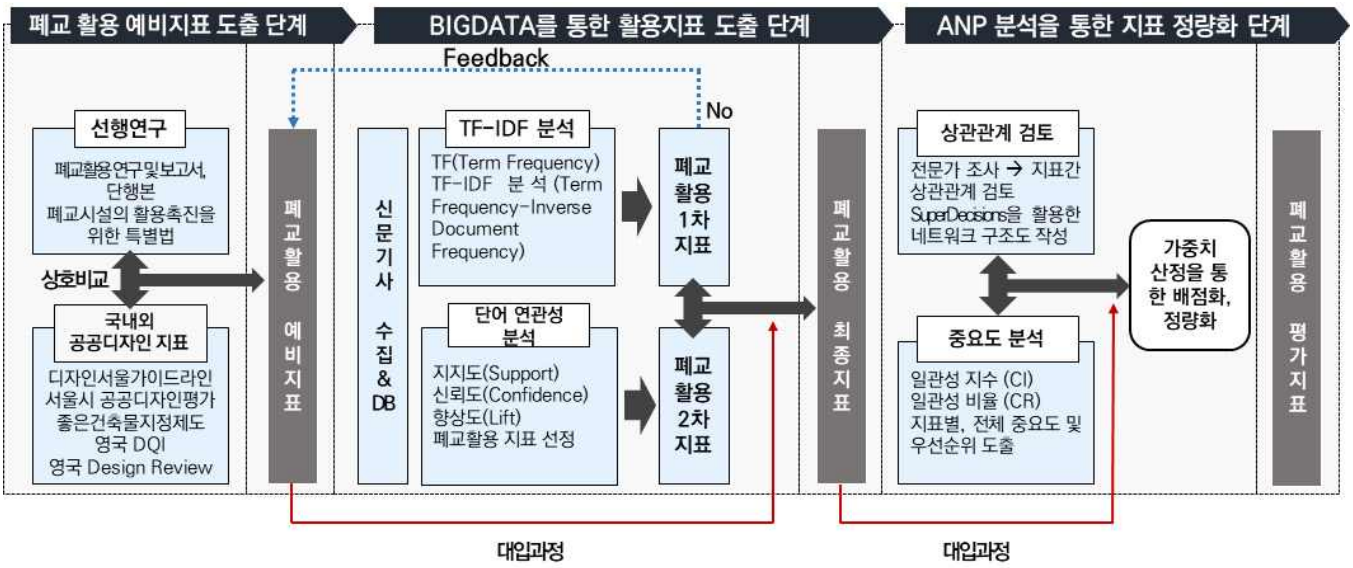

Figure1. System diagram of research

구성은 도출된 활용지표 간 상관성 검토결과를 바탕으로 5개 대분류 항목과 15개 지표 항목을 중심으로 9점 척도를 통해 쌍대비교하였다.

II. 폐교시설의 활용 예비지표 도출

II-1. 예비지표 도출 타당성 검토

본 연구에서는 폐교시설의 활용을 위한 예비 분류체계의 도출을 위해서 관련 선행연구와 국내외 공공디자인 지표를 체계화하여 상호비교 하였다. 연구범위를 공공디자인지표로 설정한 것은 폐교 활용과 관련된 선행연구는 활용 요소가 지표화되어 있지 않기 때문에 기준이 될 수 있는 지표의 설정이 필요하다. '폐교재산의 활용촉진을 위한 특별법(약칭 폐교법)' (2014.3.11. 시행) 제1조(목적)에서 '이 법은 폐교재산을 교육용시설, 사회복지시설, 소득증대시설 등의 건전한 용도로 활용하도록 촉진함으로써 평생교육 및 복지 기회를 확충하고 소득증진을 통하여 지역사회의 발전에 기여함을 목적으로 한다.'명시한 것과 같이 폐교시설을 공공을 위한 목적으로 활용하도록 유도하고 있으며 국내외 공공디자인들도 공공시설 및 건축의 공공성 확보와 품질 향상을 목적으로 하고 있으므로 목적성이 부합한다고 할 수 있다.

다음으로 폐교법과 각각의 시설 관련 법규에서 정의하는 시설 종류 외에도 실제 폐교시설의 세부활용현황의 조사[2]를 자체활용, 대부 및 임대시설로 구분하여 진행하였다. 이를 통해 활용 시설을 유형화하여 관련된 공공건축의 세부범위와의 연관성을 검토하였다. 검토 대상 폐교시설은 자체활용시설 312개, 대부 및 임대시설 681개를 대상으로 하였다. 검토 내용은 <Figure2>와 같다.

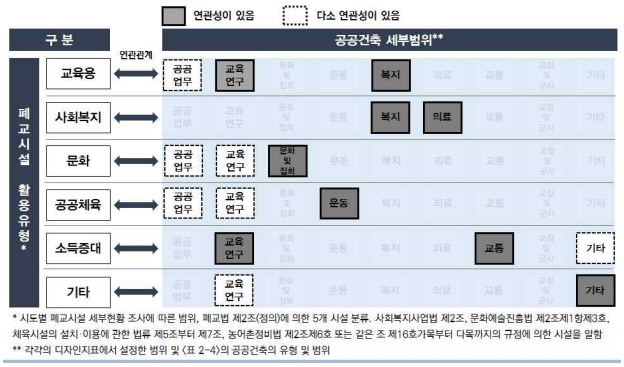

Figure2. Relationship between closed school facilities and the detailed scope of public architecture

2) 지방교육재정알리미, http://www.eduinfo.go.kr

II-2. 선행연구와 공공디자인 지표 체계화

선행연구의 검토를 통해 도출된 활용요소를 <Table1>과 같이 체계화하였다.

Table1. Systematization of pre-research utilization elements

대분류		중분류	
Code	분류	Code	요소(Element)
AR1	경제적 요인	E1	경제성
		E2	지속성
AR2	공간적 요인	E3	기능성
		E4	쾌적성
		E5	편의성
		E6	활용성
AR3	물리적 요인	E7	안전성
		E8	접근성
		E9	확장성
AR4	지역사회적 요인	E10	연계성
		E11	지역성
		E12	참여성
		E13	타당성
AR5	환경적 요인	E14	경관성
		E15	입지성
		E16	친환경성

다음으로 국내외 공공디자인 지표의 검토를 통해 도출된 지표를 <Table2>와 같이 체계화하였다.

Table2. Systematization of public design indicators in Korea and abroad

대분류		중분류	
Code	분류	Code	지표(Index)
PD1	공공성	I1	사용성 및 이용 편의성
		I2	안전성
		I3	접근성
PD2	기능성	I4	연계성
		I5	합리적 공간계획
PD3	디자인성	I6	경관 및 디자인
		I7	조화성
PD4	지역성	I8	적합성
		I9	지역사회성
		I10	지역적 파급효과
PD5	품질 성능성	I11	실내환경
		I12	유지관리
		I13	지속가능성

앞서 체계화한 선행연구의 활용요소와 공공디자인 지표를 중심으로 상호 비교를 통해 연관 정도를 다음 <Figure3>과 같이 검토하였다.

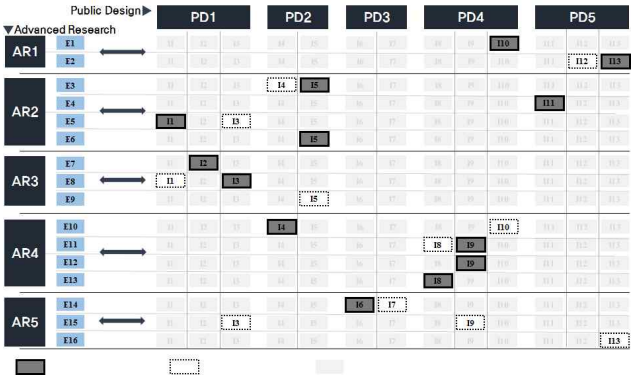

Figure3. Comparative study of pre-study utilization factors and public design indicators

선행연구 활용요소의 정의와 공공디자인 지표의 세부사항을 중심으로 상호 연관성 검토결과, E1 경제성은 I10 지역적 파급효과와 밀접한 연관이 있다. E2 지속성은 I13 지속가능성과 밀접한 연관이 있고, I12 유지관리와 다소 연관성이 있다. E3 기능성은 I5 합리적 공간계획과 밀접한 연관성이 있고, I4 연계성과 다소 연관성이 있다. E4 쾌적성은 I11 실내환경과 밀접한 연관성이 있다. E5 편의성은 I1 사용성 및 이용 편의성과 밀접한 연관성이 있고, I3 접근성과 다소 연관성이 있다. E6 활용성은 I5 합리적 공간계획과 연관성이 있다. E7 안전성은 I2 안전성과 밀접한 연관성이 있다. E8 접근성은 I3 접근성과 밀접한 연관성이 있고, I1 사용성 및 이용편의성과 다소 연관성이 있다. E9 확장성은 I5 합리적 공간계획과 다소 연관성이 있다. E10 연계성은 I4 연계성과 밀접한 연관성이 있고, I10 지역적파급효과와 다소 연관성이 있다. E11 지역성은 I9 지역사회성과 밀접한 연관성이 있고, I8 적합성과 다소 연관성이 있다. E12 참여성은 I9 지역사회성과 밀접한 연관성이 있다. E13 타당성은 I8 적합성과 밀접한 연관성이 있다. E14 경관성은 I6 경관 및 디자인과 밀접한 연관성이 있고, I7 조화성과 다소 연관성이 있다. E15 입지성은 I3 적합성과 I9 지역사회성과 다소 연관성이 있다. E16 친환경성은 I13 지속가능성과 다소 연관성이 있다.

이를 종합하여 다음 <Figure4>와 같이 폐교시설 활용 예비지표 분류 체계도를 도출하였다.

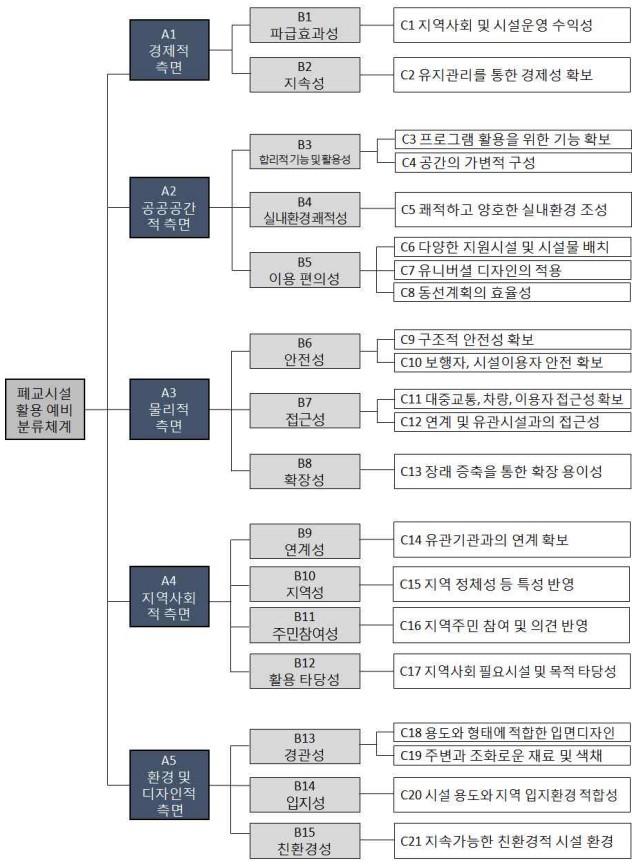

Figure4. Hierarchical model of preliminary classification system utilizing closed school facilities

III. 빅데이터를 활용한 폐교시설의 활용지표 도출

III-1. 데이터 수집 및 전처리

폐교시설 활용과 관련된 빅데이터 텍스트 마이닝 분석을 위한 자료수집을 위해 신문기사 데이터베이스 시스템인 BIGKinds[3]를 활용하였다. 자료수집의 범위는 폐교재산의 활용촉진을 위한 특별법이 개정·시행된 2014년 3월 11일을 기준으로 2019년 3월까지 폐교시설 활용에 관하여 보도된 중앙지와 지역종합지를 대상으로 하였다. 폐교 활용과 관련된 키워드가 포함된 신문기사를 수집하여 중복되는 기사와 광고성 기사를 제거하고 최종적으로 2,384개의 기사자료를 수집하였다.

수집된 기사자료는 자연어로 작성된 비정형텍스트이기 때문에 분석을 위해 자연어 처리가 필요하다. 먼저 문서형태의 자료를 품사 단위로 구분하여 필요한 데이터를 추출하는 전처리를 위해 연구 범위에서 언급하였듯이 'R'언어의 KoNLP, tm 패키지를 이용

3) BIGKinds, https://www.bigkinds.or.kr(2019년3월7일 검색)

하여 형태소 분석 및 명사형 추출을 실시하였으며, 대상 단어들의 확인과 확인된 특수기호, 일반 특수문자, 공백 및 분석에 불필요한 단어 제거 과정을 진행하였다. 코딩 과정은 다음 <Figure5>와 같다.

형태소 분석방법은 '폐교시설이 증가하고 있다.'라는 문장을 '폐교시설', '이', '증가', '하고', '있', '다'로 분석되고, 명사형 추출은 '폐교', '시설', '증가'와 같이 단어 명사로 추출하는 것이다. 또한 신문기사에 포함된 '!', '@' 등과 같은 특수기호와 문자, 기자 이름 등과 같이 분석에 불필요한 단어를 함수에 추가하여 결과값에서 제외하게 된다.

```
#단어 앞 뒤에 특수기호가 사용된 것 확인

func.check.punct <- function(x) {str_extract_all}(x, "[[:alum:]]{1,}[[:punct:]]{1}?[[:alnum:]]{1,}")}
corpus.check.punct <-lapply(corpus, func.check.punct)
table(unlist(corpus.check.punct))

#수치자료 확인

func.check.number <- function(x) {str_extract_all(x, "[[:digit:]]{1,}")}
corpus.check.number <- lapply(corpus, func.check.number)
table(unlist(corpus.check.number))

corpus.pre <- tm_map(corpus, removeNumbers)  #숫자 모두 삭제
corpus.pre <- tm_map(corpus.pre, removePunctuation) #특수 문자 제거
corpus.pre <- tm_map(corpus.pre, removeWords, words = stopwords("SMART")) #불용어(Stopwords) 제거 (SMART 목록)

#불용어/특수문자/숫자 제거된 corous(corpus.pre)의 특수기호 여부 재확인

func.check.punct <- function(x) {str_extract_all}(x, "[[:alum:]]{1,}[[:punct:]]{1}?[[:alnum:]]{1,}")}
corpus.check.punct <-lapply(corpus, func.check.punct)
table(unlist(corpus.check.punct))

#불용어/특수문자/숫자 제거된 corous(corpus.pre)의 수치 여부 재확인

func.check.number <- function(x) {str_extract_all(x, "[[:digit:]]{1,}")}
corpus.check.number <- lapply(corpus, func.check.number)
table(unlist(corpus.check.number))

#추가 단어 제거 또는 유사단어 일괄처리

func.substitute <- function(obj, old, new) {
  sub.obj <- tm_map(obj, content_transformer(function(x, pattern) gsub(pattern, new, x)), old)
}

corpus.pre <- gsub("폐교","corpus.pre")
corpus.pre <- gsub("[[:punct:]]","corpus.pre")
corpus.pre <- gsub(corpus.pre, side="right")
corpus.pre <- gsub("","corpus.pre")
corpus.pre <- gsub("^[a-zA-Z0-9]*","corpus.pre")
corpus.pre <- gsub("서남대","corpus.pre")
```

Figure5. Data pre-processing coding process using R

III-2. TF-IDF분석을 통한 1차 활용지표 도출

2장에서 도출된 예비활용지표의 검증을 위해 빈도분석을 통해 도출된 상위 키워드를 각각의 활용지표를 중심으로 군집화하여 TF분석과 TF-IDF 분석을 실시하였다. 빈도분석 결과 가장 많은 빈도수를 보인 유형은 입지성(농어촌, 도서벽지, 도심지역 등, 3,869개)으로 나타났고, 다음으로는 타당성, 주민 참여성, 파급효과성, 지역성, 이용 편의성, 안전성, 연계성, 경관성, 접근성, 지속성, 합리적 기능 및 활용성, 확장성, 친환경성, 실내환경 쾌적성 순으로 나타났다. 입지성 유형의 빈도수가 높게 나타난 것은 폐교시설의 발생 배경과 지역에 대한 언급이 자주 언급되기 때문이다. 반면에 실내환경 쾌적성 유형은 빈도수가 328개로써 상대적으로 매우 낮은 결과를 보였으며, 키워드로 도출된 쾌적, 환경은 기존에 방치된 폐교시설의 쾌적한 자연환경의 조성과 주민들의 환경 활동 등 폐교시설의 실질적인 활용과 거리가 멀기 때문이다.

전체 기사에서 빈도가 높은 단어들 중에는 단순히 발생 배경에 대한 내용이나 관용적으로 사용되는 단어들을 제외하여야 하므로 TF-IDF 분석을 통해 해당 문서에서 중요도가 높은 키워드를 도출하는 것이 필요하다. tm 패키지를 이용한 TF-IDF 분석 결과 높은 값을 나타낸 키워드 유형은 파급효과성(지역경제, 소득증대, 관광객 등, 39.57)이 가장 높게 나타났고, 다음으로는 주민참여성, 타당성, 지역성, 입지성, 안전성, 연계성, 지속성, 접근성, 합리적 기능 및 활용성, 확장성, 이용 편의성, 친환경성, 경관성, 실내환경 쾌적성 순으로 나타났다. <Table3 참조>이는 단순 빈도분석 결과와는 다른 결과를 보여준다. 여기서 TF-IDF 순위는 도출된 값에 따른 순위이며, 이 결과를 바탕으로 본 연구에서는 TF-IDF 값이 5.00 이하, 전체값의 2% 이하로 나타난 상대적으로 중요도가 낮은 경관성과, 실내환경 쾌적성 지표항목을 폐교시설의 최종 활용지표에서 제외하였다.

III-3. 키워드 연관관계 분석을 통한 2차 활용지표 도출

항목간 관련성을 파악하여 둘 이상의 항목들로 구성된 연관성 규칙을 도출하는 탐색적 자료분석 방법을 연관분석이라 한다. 연관분석을 위해 'R'프로그램의 arules 패키지와 apriori 알고리즘을 이용하였으며, 이는 사건 A가 일어났을 때, B가 일어날 확률을 나타내는 것으로써 사건 A와 B사이에 연계한 규칙을 찾는 것이다. 연관성을 측정하는 지표로는 Support, Confidence, Lift가 있다. support는 전체 기사 중 두 개의 키워드가 동시에 존재하는 기사의 비이며, confidence는 조건부확률로써 키워드 A를 포함하고 있는 기사 중 키워드 B가 포함되어 있을 확률을 나타낸다. support와 confidence는 1에 가까울수록 연관성이 높다고 본다.

Lift는 키워드 B를 연관규칙과 관계없이 관찰하는 것에 비하여 연관규칙을 알고 키워드 A를 발견한 경우에 대한 키워드 B가 관찰되는 경우 얼마나 관찰이 증가하는가를 나타내는 측도이다. 즉, 키워드 A와 B가 관련이 없다면 향상도는 1이 되며 1보다 크면 클수록 양의 연관성을 가지고, 1보다 작으면 음의 연관성을 가진다. 산출식은 다음과 같다.

$$\text{Support}(A,B) = \Pr(A \cap B) \quad <1>$$

$$\text{Confidence}(A \rightarrow B) = \frac{P_r(A \quad B)}{P_r(A)} \quad <2>$$

Table3. Results of TF-IDF analysis of preliminary utilization indicators

대분류	활용지표	텍스트 마이닝을 통해 도출된 단어	TF (Term Frequency)	TF 순위	TF-IDF	TF-IDF 순위
경제적 측면	파급 효과성	지역경제, 소득증대, 관광객, 수익성, 일자리, 관광거점, 평생교육, 문화창출, 복지증대, 지역발전	3,423	4	39.57	1
	지속성	유지관리, 보존관리, 효율성, 지속가능, 임대기간	873	11	26.45	8
공공 공간적 측면	합리적 기능 및 활용성	활용성, 기능성, 가변성, 활용도, 리모델링, 공간구성, 공공시설, 공간효율, 공간설계	723	12	20.18	10
	실내환경 쾌적성	마을경관, 미관, 흉물	98	15	3.71	15
	이용 편의성	편의시설, 편의공간, 휴식공간, 편리성	2,124	6	18.03	12
물리적 측면	안전성	노후화, 안전문제, 안전진단, 안전사고, 안전대책, 안전관리	1,583	7	32.21	6
	접근성	접근성, 접근로, 교통, 대중교통, 거리	989	10	25.39	9
	확장성	증축, 시설확장, 확장성	454	13	18.35	11
지역 사회적 측면	연계성	유관기관, 연계활용, 네트워크	1,371	8	28.78	7
	지역성	지역문화, 역사도시, 관광지, 특성화, 정체성, 역사성, 관광자원	2,874	5	33.22	4
	주민 참여성	지역주민, 주민투표, 공동체, 주민자치위원회, 의견수렴	3,612	3	38.84	2
	활용 타당성	사업비, 지원사업, 도시재생, 투자성, 활용률, 활용가치, 미활용, 방치, 사업계획서, 경제성	3,653	2	38.13	3
환경 및 디자인적 측면	경관성	마을경관, 미관, 흉물	1,192	9	4.78	14
	입지성	농어촌, 도서벽지, 도심지역, 산간벽지, 자연환경, 보호구역, 구도심	3,869	1	32.74	5
	친환경성	자연생태, 친환경, 생태공간	203	14	15.04	13

$$\text{Lift}(A \rightarrow B) = \frac{P_r(A \quad B)}{P_r(A)P(B)} = \frac{P_r(B|A)}{P(B)} = \frac{Confidence}{P(B)} \qquad <3>$$

이 산출식에 근거한 알고리즘을 바탕으로 도출된 기사자료를 support를 0.01, confidence를 0.3 이상인 규칙만을 추출하였으며, 그 결과 도출된 연관규칙 중 상위 Lift 값을 가지는 규칙 15개는 다음 <Table4>와 같다.

표를 예를 들면 '교육청'과 '폐교'가 동시에 발견될 확률인 support는 0.02로 작게 나타났지만, 선행 키워드 '교육청'이 포함된 기사에서 후행 키워드 '폐교'가 발견될 확률인 confidence는 0.8로 높게 나타났다. 또한 키워드 간 연관성을 의미하는 lift의 경우는 28.7로 동시에 발견될 확률의 비율이 매우 높다는 것을 의미한다.

Table4. Keyword association analysis

선행 키워드 (A)	후행 키워드 (B)	지지도 (Support)	신뢰도 (Confide-nce)	향상도 (Lift)
교육청	폐교	0.02	0.8	28.7
문화	예술	0.01	0.7	26.5
폐교	대학	0.02	0.5	25.8
폐교부지	임대료	0.01	0.8	23.2
지자체	활용방안	0.02	0.8	21.9
재산가치	건물 (대지)	0.01	0.4	20.4
학생수	폐교위기	0.03	0.8	19.2
저출산	고령화	0.01	0.9	18.7
활용	귀농귀촌	0.01	0.3	18.3
예술	체험활동	0.01	0.4	18.2
지역	활성화	0.01	0.6	15.3
리모델링	캠핑장	0.01	0.9	11.9
토지	공시지가	0.01	0.9	10.3
인구감소	폐쇄	0.01	0.6	10.1
수요자	유동인구	0.01	0.4	9.7

도출된 15개의 규칙 중 중복되는 내용과 폐교 활용지표 선정과 관련이 없는 키워드(예를 들면 '문화', '예술' 의 경우 폐교시설의 특정 활용 목적을 지향하므로 배제함)를 제외하고, 추가적인 활용지표로 적용될 수 있는 키워드 '폐교부지', '임대료', '재산가치', '건물(대지)', '학생수', '저출산', '고령화', '토지', '공시지가', '수요자', '유동인구'를 중심으로 지표항목을 추가였다.

III-4. 폐교시설 활용 최종지표 도출

폐교시설 활용 예비지표를 중심으로 TF-IDF 분석과 키워드 상관관계 분석을 통해 1, 2차 활용지표를 도출하였으며, 이를 종합하여 폐교시설 최종 활용지표를 도출하였다.

최종적으로 도출된 폐교시설의 활용지표의 대분류는 경제적, 공공 공간적, 물리적, 사업계획 수립, 시설 및 지역환경으로 구분된다. 기존 예비지표와 변경된 내용은 기존에 환경 및 디자인적 측면 항목이 삭제되면서 '친환경성'지표가 공공 공간적 측면으로 이동되었고, 키워드 연관관계 분석을 통해 시설 및 지역환경 측면 항목이 추가되면서 기존의 지역사회적 항목의 '입지성'지표가 해당 항목으로 이동되었다. 또한 지역사회적 항목의 지표 구성이 폐교활용 사업계획서 수립 시 필요한 지표들로 구성되어 항목명이 사업계획 수립측면으로 수정되었다.

추가된 시설 및 지역환경 측면의 '인구분포특성'은 폐교시설이 위치한 지역의 인구 분포 특성에 대한 내용으로써 연령별 인구분포, 인구 유동성 데이터를 기준으로 폐교시설의 활용 시설 유형이 결정될 수 있고, 폐교 활용시설의 수요를 예측할 수 있는 지표이다. 또한 '시설 및 토지 특성'은 폐교시설의 공시지가와 대장가격, 임대료 등을 기준으로 재정적인 가치를 평가함으로써 폐교 활용 여부를 결정하는 중요한 지표로 작용한다. 폐교시설의 최종 활용지표를 계층 구조 모형화 하면 다음 <Figure 6>과 같다.

IV. 폐교시설 활용지표의 정량화

본 장에서는 3장에서 도출된 폐교시설의 최종 활용지표를 기준으로 지표 간 상관관계 검토와 중요도 분석, 가중치 설정을 통해 배점화·정량화를 진행하였다. 가중치를 부여하는 방식으로는 일반적으로

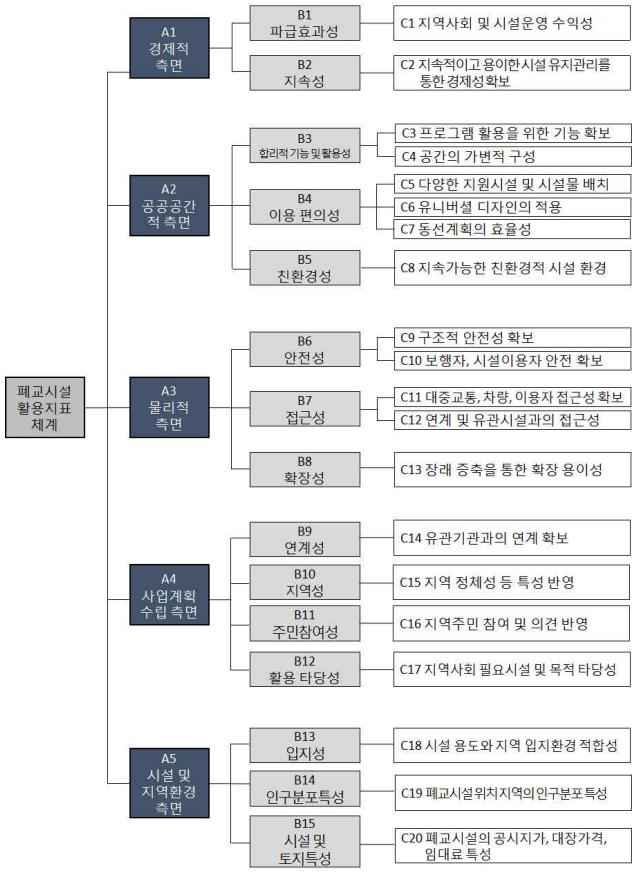

Figure6. A Hierarchical model for the utilization index of closed school facilities

객관화된 수치가 있을 경우 회귀분석, 요인분석, 상관관계 분석 등의 통계적 기법을 바탕으로 가중치를 산정할 수 있으나, 지표 간의 정량적인 분석의 기준이 없이 상대적인 비교에 따라 가중치를 설정해야 하므로 지표 간 상관관계 검토를 통해 네트워크 구조를 구성하여 상대적 중요도를 도출할 수 있는 ANP분석을 통하여 활용지표의 가중치를 산출하였다. 이를 통해 폐교 활용 시 객관적으로 판단할 수 있는 기준을 마련하여 합리적인 폐교 활용 의사결정에 활용되고자 한다. 지표의 정량화 단계는 <Figure7>과 같다.

IV-1. 지표 간 상관관계 검토

활용지표의 ANP 중요도 분석 및 가중치 설정을 위한 사전 단계로 지표간 상관관계의 검토가 필요하다. 상관관계의 검토를 위해 관련 전문가 5인(건축 3인, 디자인2인)을 대상으로 본 연구의 취지와 목적을 설명한 후 지표간 상관성이 있다고 생각되는 항목에 표기하는 방법으로 설문조사를 실시하였다.

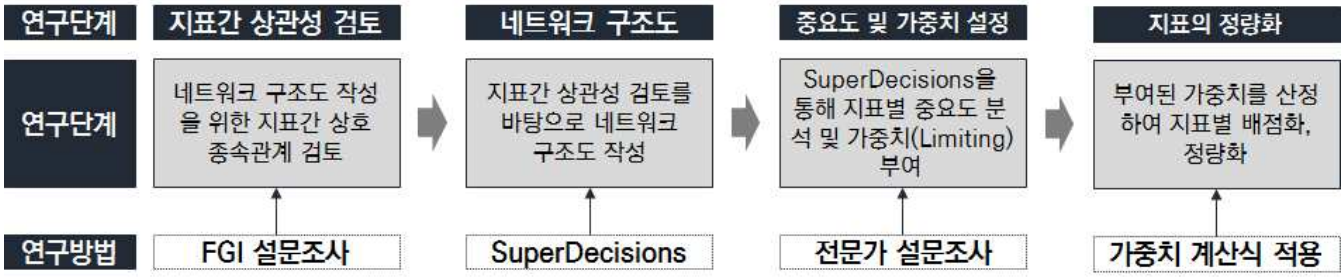

Figure7. Quantification process of indicators through ANP analysis

상관관계 규명은 좌우행렬에 상관관계가 있다고 생각되는 항목에 표기하여 70% 이상이 동일하게 대답한 결과를 상관관계가 있는 것으로 판단하였다. 분석결과는 <Table5>와 같으며, 정방행렬 형식으로 행렬의 좌측요소가 상단 요소에 영향을 주는 형태이다. 해당 셀에 '■'표시가 된 것은 셀 좌측의 요소가 셀 상단의 요소에 영향을 준다는 것을 의미하며, '■'표시가 없는 셀은 해당 셀의 좌측 요소가 셀 상단 요소에 영향을 미치지 않음을 의미한다.

Table5. Correlation between indicators results

구분		경제적		공공 공간적			물리적			사업계획 수립				시설 및 지역 환경		
		a	b	c	d	e	f	g	h	I	j	k	l	m	n	o
경제적	a	–	■						■			■	■			
	b	■	–									■	■			
	c		■	–	■											
공공공간적	d			■	–			■								
	e		■			–										
	f		■				–					■				
물리적	g							–		■		■			■	
	h	■							–	■						
	i	■	■						■	–		■				
사업계획수립	j	■									–	■	■			
	k	■	■							■		–	■			
	l		■									■	–			
시설및지역환경	m	■						■	■	■				–	■	
	n	■	■		■					■		■			–	
	o		■							■			■			–

a.파급효과성 b.지속성 c.합리적 기능 및 활용성 d.이용 편의성 e.친환경성 f.안전성 g.접근성 h.확장성 I.연계성 j.지역성 k.주민참여성 l.활용타당성 m.입지성 n.인구분포특성 o.시설 및 토지특성

상관관계 분석을 바탕으로 네트워크 분석적 의사결정 시뮬레이션을 구현할 수 있는 SuperDecisions 2.8을 사용하여 네트워크 구조도를 작성하였다. <Figure8 참조> 그림에서 직선 표시의 화살표는 클러스터 간의 상관관계가 있음을 의미하며 루프 표시의 화살표는 각 평가요소별로 내향종속성을 가진다는 것을 의미한다.

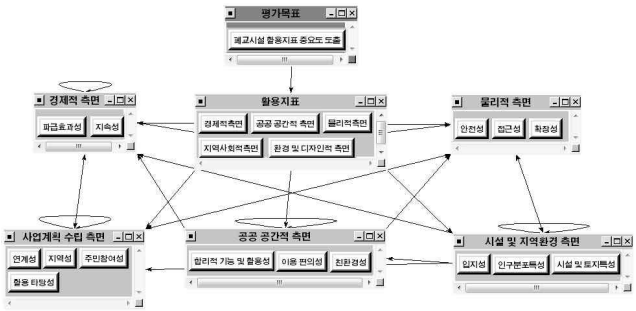

Figure8. Network model based on cross-index correlation review

IV-2. ANP 분석을 위한 설문지 작성 및 수집

설문조사는 관련 전문가 및 실무자(건축, 도시, 관련 공무원) 100명을 대상으로 설문조사를 실시하였다. 표본 수는 100부를 배포하여 48부를 회수하였으며, 회수된 표본의 신뢰성을 검증하기 위해 일관성 검증을 실시하였다.

일관성 지수(CI) = (λmax - n) / (n - 1) <4>
일관성 비율(CR) = (CI / RI) x 100% <5>

여기서, CI가 작을수록(0에 가까울수록)일관성이 양호하며, 완전 일관성일 때 λmax=n 가 되어 CI=0이 된다. 속성개수 n이 증가할수록 CI값은 점점 커지기 때문에 이를 보완할 일관성 비율 CR이 이용된다. CR은 CI를 랜덤 인덱스 RI(Random Index)로 나눈 값으로써, CR이 작을수록 일관성이 양호한 것

이며, 일관성 비율(Consistency ratio: CR)의 판단기준인 0.1보다 작은 경우 일관성이 있다고 판단한다. 판단 기준 0.1 보다 작은 유효 표본은 42부로 나타났다. 그 외의 표본은 분석에서 제외하였다. 분야별 응답자와 경력별 응답자 현황은 다음과 같다.

Table6. Current status of respondents by sectoral

구분	건축분야	도시 분야	관련 공무원	계
응답자수 (인)	32	10	6	48
구성비 (%)	66.7	20.8	12.5	100.0

Table7. Current status of respondents by experience

구분	5년 이하	6~10년	11~15년	16~20년	21년 이상	계
응답자수 (인)	3	8	17	13	7	48
구성비 (%)	6.3	16.7	35.3	27.1	14.6	100

설문지의 구성은 도출된 활용지표 간 상관성 검토결과를 바탕으로 5개 대분류 항목과 15개 지표 항목을 중심으로 9점 척도를 통해 쌍대비교하였다. 예를 들어 경제적 측면과 공공공간적, 물리적, 사업계획 수립, 시설 및 지역환경 측면을 각각 비교하고, 공공 공간적 측면과 물리적, 사업계획 수립, 시설 및 지역환경 측면을 각각 비교하는 방식이다. 설문 문항 예시는 <Figure9>와 같다.

1) 폐교시설 활용지표의 상위 부문인 **경제적 측면, 공공 공간적 측면, 물리적 측면, 사업계획 수립 측면, 시설 및 토지환경 측면**을 상호 비교하여 어느 정도로 중요한지를 표기(√)하여 주시기 바랍니다.

판단기준 (A)	극히 9		매우 7		중요 5		약간 3		비슷 1		약간 3		중요 5		매우 7		극히 9	판단기준 (B)
경제적 측면																		공공 공간적 측면
경제적 측면																		물리적 측면
경제적 측면																		사업계획 수립 측면
경제적 측면																		시설 및 지역환경 측면
공공 공간적 측면																		물리적 측면
공공 공간적 측면																		사업계획 수립 측면
공공 공간적 측면																		시설 및 지역환경 측면
물리적 측면																		사업계획 수립 측면
물리적 측면																		시설 및 지역환경 측면
사업계획 수립 측면																		시설 및 지역환경 측면

Figure9. Example of survey items for ANP analysis

IV-3. 활용지표의 중요도 분석 및 정량화

활용지표를 정량화하기 위해 지표의 전체적인 중요도를 <Table8>과 같이 분석하였다. 지표별 중요도(Nomalized by cluster)값은 각 상위항목 별로 절대값 1을 기준으로 배분된 중요도 값이며, <Figure 9> 설문조사를 통해 도출된 것이다.

전체 중요도(Limiting)값은 평가지표 전체를 절대값 1을 기준으로 배분된 중요도 값을 의미한다. 가중치 부여값은 전체 중요도를 100점 환산하였다. 이를 통해 도출 된 가중치 부여 값을 배점이 용이하도록 소수점을 조정하였다. 활용지표의 정량화 결과, 분야별 배점은 전체 100점을 기준으로 경제적 측면 20점, 공공 공간적 측면 15점, 물리적 측면 21점, 사업계획 수립 측면 25점, 시설 및 지역환경 측면이 19점으로 배분되었다. 경제적 측면의 파급효과성과 지속성이 각각 10점, 공공 공간적 측면의 합리적 기능 및 활용성 10점, 이용 편의성 3점, 친환경성 2점, 물리적 측면의 안전성 7점, 접근성 11점, 확장성 3점, 사업계획 수립 측면의 연계성 4점, 지역성 3점, 주민 참여성 11점, 활용 타당성 6점, 시설 및 지역환경 측면의 입지성 7점, 인구 분포특성 3점, 시설 및 토지특성이 9점으로 배분되었다.

V. 결 론

본 연구는 지속적으로 증가가 예상되는 폐교시설의 효율적인 활용을 위하여 활용지표의 개발을 통해 폐교 활용 시 객관적인 의사결정을 위한 목적을 가진다. 연구 단계는 크게 폐교 활용 예비지표 도출, 빅데이터를 활용한 최종지표 도출, 지표의 정량화 단계로 구분되며 결과를 요약하면 다음과 같다.

첫째, 폐교시설의 활용 예비지표 도출단계에서 선행연구의 활용요소와 국내외 공공디자인 지표 검토를 통해 유사한 항목간 통합하여 중복성을 확인한 후 체계화 하였다. 체계화된 지표의 상호 비교를 통

https://doi.prg/10.7743/kisee.2019.18.2.001

Table8. Quantification of utilization indicators

구분	지표	중요도 (Nomalized by cluster)	전체 중요도 (Limiting)	가중치 부여값	평가배점
5개 지표항목	경제적 측면	0.201	-	20.1	**20**
	공공 공간적 측면	0.149	-	14.9	**15**
	물리적 측면	0.211	-	21.1	**21**
	사업계획 수립 측면	0.245	-	24.5	**25**
	시설 및 지역환경 측면	0.194	-	19.4	**19**
	소계			100	**100**
경제적 측면	파급 효과성	0.624	0.105	10.5	10
	지속성	0.376	0.096	9.6	10
	소계			20.1	**20**
공공 공간적 측면	합리적 기능 및 활용성	0.433	0.095	9.5	10
	이용 편의성	0.352	0.029	2.9	3
	친환경성	0.215	0.025	2.5	2
	소계			14.9	**15**
물리적 측면	안전성	0.383	0.073	7.3	7
	접근성	0.485	0.112	11.2	11
	확장성	0.132	0.026	2.6	3
	소계			21.1	**21**
사업계획 수립측면	연계성	0.187	0.043	4.3	4
	지역성	0.212	0.034	3.4	3
	주민 참여성	0.318	0.109	10.9	11
	활용 타당성	0.283	0.059	5.9	6
	소계			24.5	**25**
시설 및 지역환경 측면	입지성	0.332	0.071	7.1	7
	인구 분포특성	0.281	0.031	3.1	3
	시설 및 토지 특성	0.387	0.092	9.2	9
	소계			19.4	**19**
합 계				**100**	**100**

해 연관성을 검토하였으며 이를 바탕으로 활용 예비 분류체계를 도출하였다.

둘째, 빅데이터를 활용한 텍스트마이닝 분석을 위해 데이터 수집 및 전처리 과정을 실시하였다. 폐교시설 활용과 관련한 중앙지와 지역종합지를 대상으로 최종적으로 2,384개의 자료를 확보하였으며, R 프로그램의 패키지를 활용한 불용어 및 공백제거 등 데이터 전처리를 통해 수집된 기사자료에서 필요한 단어 및 명사를 추출하였다. 전처리된 기사에 대하여 빈도(Term frequency)분석과 TF-IDF(Term frequency-Inverse document frequency) 분석을 실시하여 1차 활용지표를 도출하였으며, 빅데이터를 활용한 폐교시설 활용지표의 추가를 위해 키워드 연관관계 분석을 실시하여 2차 활용지표를 도출하였다.

셋째, 폐교시설 활용 예비지표를 중심으로 TF-IDF 분석과 키워드 상관관계 분석을 통해 1, 2차 활용지표를 도출하였으며, 이를 종합하여 폐교시설 최종 활용지표를 도출하였다. 최종적으로 도출된 폐교시설의 활용지표의 대분류는 경제적, 공공 공간적, 물리적, 사업계획 수립, 시설 및 지역환경으로 구분된다.

넷째, ANP 분석을 활용하여 활용지표를 정량화하였다. 지표별 전체 중요도 분석을 통해 가중치 부여 계산식에 근거하여 가중치를 부여하였으며, 이를 배점이 용이하도록 소수점을 조정하여 분야별 배점은 전체 100점을 기준으로 경제적 측면 20점, 공공 공간적 측면 15점, 물리적 측면 21점, 사업계획 수립 측면 25점, 시설 및 지역환경 측면이 19점으로 배분되었다.

본 연구는 지금까지 폐교시설의 활용을 위한 계획 및 연구에 있어서 시도되지 않았던 빅데이터 분석기법을 적용한 것에 그 의의가 있다. 그러나 폐교시설의 활용을 위한 통합적 관점에서 접근하였기 때문에 다양한 폐교시설의 지역적 특성과 시설 특성이 배제된 한계가 있다. 따라서 향후 정량화된 지표를 바탕으로 실제 시설에 적용하여 검증하는 단계가 필요하다.

국문초록

본 연구는 지속적으로 증가가 예상되는 폐교시설의 효율적인 활용을 위하여 활용지표의 개발을 통해 폐교 활용 시 객관적인 의사결정을 위한 목적을 가진다. 연구 단계는 크게 폐교 활용 예비지표 도출, 빅데이터를 활용한 최종지표 도출, 지표의 정량화 단계로 구분하여 진행하였으며, 최종적으로 지표를 정량화함으로써 객관화하였다. 향후 지표를 기준으로 시설에 적용 및 검증하고자 한다. 본 연구는 지금까지 폐교시설의 활용을 위한 계획 및 연구에 있어서 시도되지 않았던 빅데이터 분석기법을 적용한 것에 그 의의가 있다.

참고문헌

1. 지방교육재정알리미, http:/www.eduinfo.go.kr
2. BIGKinds. https://www.bigkinds.or.kr
3. 김동석(2017). 빅데이터 분석 기법을 활용한 도서관발전종합계획 동향 분석. 건국대학교 박사학위논문
4. 김재영, 이종국(2019). 폐교시설의 활용모형 개발을 위한 예비 분류체계 도출 연구. 청소년시설환경, 17(1). 131-141.
5. 김상호, 김영현, 이민우(2009). 공공건축의 디자인 향상을 위한 디자인품질지표 개발 연구. 안양, 건축도시공간연구소.
6. 유예림(2017). 빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석. 서울대학교 박사학위논문
7. 정영훈(2006). ANP를 적용한 정보시스템 아웃소싱 결정모델. 충남대학교 석사학위논문
8. Gantz. J., Reinsel. D(2011). Extracting value from chaos. IDC Iview, 1142, 1-12.

(논문투고일 : 2019.05.20, 심사완료일 : 2019.06.13, 게재확정일 : 2019.06.19.)