

클러스터링 기법을 활용한 해외건설 필요정보 우선순위 수요 조사 평가*

Priority Demand Assessment for Overseas Construction Information Using Clustering Method

최 원 영 (Wonyoung Choi)**

곽 승 진 (Seing-Jin Kwak)***

초 록

국내 건설시장의 침체가 예상되는 상황에서 지속적으로 국내 중소엔지니어링 기업의 해외시장으로의 진출을 지원하기 위해 해외건설엔지니어링 정보시스템(OVICE)이 운영되고 있다. 이에 본 연구에서는 기존 연구를 통해 수행된 전문가 설문조사를 통한 필요정보 우선순위와 정보시스템 사용자 통계를 비교하여 정보시스템의 정보제공 방향성을 제시하여 정보서비스 품질을 높이고자 하였다. 정보시스템의 이용 통계 분석에 있어서, 우선순위를 매기기 힘든 통계분석의 효율성을 높이기 위해 K-means Clustering 기법을 분석에 활용하였다. 그 결과 기존 설문결과와 정보시스템 이용 통계의 차이를 분석하여 정보시스템의 정보제공에서의 보완점과 함께 설문조사 과정에서 부각되지 않았던 중요한 콘텐츠를 찾아낼 수 있었다.

ABSTRACT

In a situation when domestic construction market is expected to be stagnant, Overseas Information System for Construction Engineering (OVICE) is operated to support the construction SMEs that advance to the global market. In this study, we aimed to improve the quality of information service by providing direction of information provision, by comparing expert questionnaire with information system user statistics. For statistical analysis of information systems, to improve the efficiency of statistical analysis that is difficult to prioritize, K-means clustering is used for more efficient analysis. As a result, analyzing the difference between the survey results and the information system statistics, we were able to identify improvement point of information provision in the system and important contents that were not highlighted during the survey.

키워드: 해외건설엔지니어링 정보시스템(OVICE), 해외건설정보 분류체계, K-평균 군집화, 실루엣 기법, 정보수요조사
Overseas Information System for Construction Engineering (OVICE), Overseas Construction Information Classification, K-means Clustering, Silhouette Method, Information Demand Survey

* 본 연구는 국토교통부 건설기술연구사업의 연구비지원(18SCIP-C079445-05)에 의해 수행되었음.

** 한국건설기술연구원 연구원(wonyoung@kict.re.kr) (제1저자)

*** 충남대학교 문헌정보학과 교수(sjkwak@cnu.ac.kr) (교신저자)

논문접수일자 : 2018년 11월 10일 논문심사일자 : 2018년 11월 27일 게재확정일자 : 2018년 11월 28일
한국비블리아학회지, 29(4) : 57-68, 2018. [http://dx.doi.org/10.14699/kbiblia.2018.29.4.057]

1. 서론

1.1 연구의 목적 및 방법

국내 건설시장은 2010년을 전후하여 장기침체가 지속되어 건설물량이 지속적으로 축소되고 건설업체의 수익성이 약화되어 경영난이 가중되고 있다. 또한 2014년 107.4조원에서 2017년 160.4조원으로 3년 동안 60조정도 일시적으로 수주액이 증가하는 경향을 보이나, 이는 주로 주택시장에 의존하고 있는 것으로(2008년 대비 토목부분 수주액 성장률 2%, 주택부분 성장률 53%) 향후 부동산 정책/시장상황에 따라 언제든지 급변할 위험성을 내포하고 있다.

이와 반면 세계건설시장은 연평균 5%이상 지속적인 성장이 이루어지고 있으며 이러한 성장세는 앞으로도 지속될 것으로 예상되고 있다. 하지만 해외건설시장의 지속적인 성장세에도 불구하고 국내기업의 해외건설실적은 2014년 660억 달러에서 최근 3년간 지속적으로 감소하여 2017년에는 290억 달러에 이르고 있다. 이는 중동지역의 발주물량 감소, 수주 내실화를 위한 보수적 입찰전략 견지, 중국의 저가수주 공세 및 선진국

형 수주구조로의 체질개선 지연에 따른 수주경쟁력 약화 등이 원인으로 추후 국내건설경기 침체에 대비해 해외건설 진출을 위한 대책마련이 시급한 상황이다.

특히 건설엔지니어링 분야는 세계 주요건설시장에 선진 각국 기업이 선점하고 있으며, 우리나라 기술력도 열세에 있어 중소기업의 경우 해외진출을 위한 경쟁력은 더욱 떨어지는 상황이다. 이러한 상황에서 국토교통부는 2016년부터 해외건설엔지니어링 정보시스템(이하 OVICE)을 개발/운영하면서, 프로젝트 발주, 입찰, 계약 등 현지화된 각종 기술정보를 제공하여 국내 중소기업의 해외경쟁력 강화를 도모하고 있다. 국내 건설시장의 연도별 총 수주액은 <표 1>과 같으며, 연도별 해외건설시장 추이 및 국내 수주액은 <표 2>와 같다.

이에 본 연구에서는 실제 OVICE 사용자의 이용통계 및 해외건설정보 수요를 조사·비교하여 정보시스템의 정보제공 방향성에 대한 판단을 위한 기초자료로 활용하여 국내 기업들의 해외경쟁력 강화를 위한 정보제공의 질을 높여 국내 건설 및 건설엔지니어링 기업의 해외진출에 기여하고자 한다.

<표 1> 연도별 국내건설시장 총 수주액(조원)

| 구분 \ 연도 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | |
|---------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|
| 토목부분 | 41.3 | 54.1 | 41.4 | 38.8 | 35.7 | 29.9 | 32.7 | 45.5 | 38.2 | 42.2 | |
| 건축 부문 | 주거 | 44.7 | 39.1 | 31.6 | 38.7 | 34.3 | 29.3 | 41.1 | 67.7 | 75.9 | 68.8 |
| | 비주거 | 34.2 | 25.5 | 30.2 | 33.2 | 31.5 | 32.1 | 33.7 | 44.8 | 50.7 | 49.5 |
| | 계 | 78.8 | 64.6 | 61.8 | 71.9 | 65.8 | 61.4 | 74.7 | 112.5 | 126.7 | 118.2 |
| 국내시장총액 | 120.0 | 118.8 | 103.1 | 110.7 | 101.4 | 91.3 | 107.4 | 158.0 | 165.0 | 160.4 | |

출처: 대한건설협회

〈표 2〉 연도별 해외건설시장 추이 및 국내 수주액(억\$)

| 구분 \ 연도 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---------|------|------|------|------|------|------|------|------|------|------|
| 해외건설성장률 | 7.2 | -6 | 8.6 | 12.7 | 4.4 | 5.8 | 4.9 | 9.5 | -4.7 | 10 |
| 해외수주액 | 476 | 491 | 716 | 591 | 649 | 652 | 660 | 461 | 282 | 290 |

출처: IHS global Insight, 해외건설협회

1.2 선행 연구

해외진출을 위한 건설분야 기업들의 해외건설정보 수요 분석과 관한 연구는 많이 이루어지지 않았으나 다음과 같이 일부 관련 연구가 진행되어 왔다.

박환표 등(2008)은 건설기업의 해외진출 활성화를 위해 업계 요구사항을 분석하여, 해외시장 정보제공의 중요성을 제기하였고, 정부의 해외건설 지원체계의 개선방안으로 해외건설 수주정보 지원 및 해외사업의 데이터베이스 지원을 제안하였다. 윤승희 등(2015)은 중소건설업체에 초점을 맞추어 설문 등을 통해 해외사업 수행 시 자체적인 문제점 및 부족 역량을 도출하고 부족역량에 대한 인과관계를 도출하여 각 문제점의 해결방안을 제시하였다. 권오철, 조정근(2017)은 해외건설 엔지니어링 정보시스템 구축을 위해 해외건설 업무 수행단계별로 필요한 정보수요 조사 및 필요정보에 대한 우선순위 도출을 수행하여, 사업정보, 국가별 건설정보, 사전조사 및 타당성조사, 계약정보 등의 분야에 대한 필요정보 우선가 높음을 확인하였다. 이와 같이 앞선 해외건설정보 수요 분석에 관한 연구는 대부분 설문조사에 기초한 연구가 대부분으로 실제 이용자 통계를 이용한 분석은 없는 것으로 볼 때, 본 연구에 상당한 의미가 있다고 볼 수 있다.

본 연구에서 사용하고자 하는 K-Means Clustering 기법은 오래전부터 연구가 시작되어 발전되어 왔다. MacQueen(1967)에 의해 용어에 대한 정의가 이루어졌으며, 무게중심을 사용한 군집화 방법의 알고리즘을 제시하였다. 하지만 국내 건설 분야에서 이러한 군집화 기법을 이용한 연구는 많지 않았다. 백승원 등(2018)은 프로젝트 진행과정에서 제기된 이슈를 도취하기 위해 연관 단어등을 대상으로 K-Means Clustering 기법을 이용하여 키워드 분석을 수행하였다. 이지섭 등(2018)은 공공 건설사업에서 갈등을 유발하는 요인들을 규명하기 위해 군집화 기법을 사용하여 갈등 사례를 평가하는 연구를 수행하였다. 따라서 군집화 기법을 사용해서 해외건설분야의 수요분석을 수행하고자 하는 본 연구에 의미를 부여할 수 있다.

반면에, 이러한 군집화 기법에서의 가장 큰 문제점은 적절한 군집의 개수를 찾는 방법이다. 군집의 수에 따라 구집의 분류 결과가 극명하게 달라지며, 좋지 못한 결과를 보여줄 가능성이 있기 때문에 적절한 k값을 찾는 방법이 매우 중요하다. 군집의 개수를 정하는 간단한 방법으로 Martin and Maes(1979)는 데이터의 개수의 절반의 제곱근의 근사값으로 군집의 개수를 정하는 방법을 제시하였다. Ketchen and Shook(1996)는 군집의 수를 순차적으로 늘려가면서 군집내의 오차제곱의 합을 계산하여 이전클러스터에

비해 변화량이 완만해 지는 지점인 elbow point 를 찾아 적절한 군집의 개수를 찾아가는 방법을 제시하였다. Rousseeuw(1987)은 군집내 데이터의 응집도, 클러스터간의 분리도를 수치적으로 나타내 군집 분류의 품질을 정량적으로 제시할 수 있는 실루엣 기법을 제시하였다. 이에 본 연구에서는 최적의 군집개수를 찾기 위해 정량적으로 수치를 비교할 수 있는 실루엣 기법을 사용하였다.

2. 연구방법

2.1 해외건설정보 분류체계

한국건설기술연구원(2018)에서는 OVICE의 장기적인 운영을 위한 중장기 운영전략계획을 수

립하면서, 해외건설전문가 수요조사를 통해 EPC 개념을 포함한 해외건설정보 분류체계를 수립하고, 정보서비스의 효율적인 정보제공을 위해 통합/수정을 거쳐 <표 3>과 같이 7개 대분류, 32개 분류체계를 운영하고 있다.

또한 OVICE에서는 국내 기업이 해외투자 개발사업 진출이 용이하다고 판단되는 국가들을 거점국가로 정하여 매년 거점국가를 추가하여 정보를 제공하고 있다. 현재까지 구축·운영 중인 거점국가는 총 9개로 현황은 <표 4>와 같다.

본 연구에서는 분류체계별 필요정보 우선순위를 선정함에 있어 국가별 특성에 따른 수요의 차이를 분석하기 위해 위의 총 9개 거점국가와 국가분류가 되지 않은 국가 미분류로 총 10개의 국가를 나누어 총 320개의 분류체계를 가지고 분석을 수행하였다.

<표 3> 해외건설정보 분류체계

| 대분류 | 중분류 | 소분류 | 대분류 | 중분류 | 소분류 | | |
|----------|--------------|-----------------|-----------|-----------|--------------|-----------------|----------------|
| 영업 및 기획 | 일반정보 | 일반시장동향 | 조사 및 설계 | 기본설계 | 기본설계/계획 도서정보 | | |
| | | 사회문화정치경제 | | | 인허가정보 | | |
| | 법령 및 제도 | 일반 법령 및 제도 | 구매조달 | 구매조달 | 구매조달 | 프로젝트별 실시설계 도서정보 | |
| | | 건설 법령 및 제도 | | | | | |
| | 건설정보 | 사업개발계획 | 공사관리 | 공사관리 | 공사관리 | 일반관리 | |
| | | 건설인프라 현황 및 환경정보 | | | | 발주처관리 | |
| 건설시장동향 | | 공정 및 원가관리 | | | | | |
| | 기술트렌드 | | | | 품질안전보건환경관리 | | |
| 입찰 및 계약 | 입찰 | 입찰관련 정보 | 운영 및 유지관리 | 운영 및 유지관리 | 협력업체 관리 | | |
| | | 입찰관련 문서 | | | 운영 및 유지관리 | | |
| | | 보증 및 보험 | | | 운영 및 유지관리 | | |
| | 견적 | 현장조사 | 공통일반 | 공통일반 | 클레임 정보 | 분쟁과 협의 및 클레임자료 | |
| | | 견적 및 물가자료 | | | 업체정보 | 업체정보 | |
| | 계약 | 계약 관리 정보 | | | 코드 및 기준 | 코드 및 기준 | 지역, 국가별 리스크 현황 |
| 계약 관리 정보 | | 리스크 관리 | | | 리스크 관리 | 지역, 국가별 리스크 현황 | 성공 및 실패사례 |
| 조사 및 설계 | 사전조사 및 타당성조사 | 지형/지질/기후정보 | 기타일반 | 기타일반 | 기타일반 | | |
| | | 조사보고서 | | | 기타일반 | | |

〈표 4〉 해외건설엔지니어링 정보시스템 해외진출 거점국가

| 지역구분 | | 거점국가 |
|-------|-------|----------------------|
| 아시아 | 동남아시아 | 베트남, 인도네시아, 필리핀, 미얀마 |
| | 남아시아 | 방글라데시 |
| 중동 | | 터키 |
| 아프리카 | | 케냐 |
| 북아메리카 | | 미국 |
| 남아메리카 | | 페루 |

2.2 OVICE 사용자 접속 통계

OVICE는 국내 중소 건설엔지니어링 기업의 해외경쟁력 강화를 위해 2016년 8월부터 서비스를 시작하여, 총 9개 거점국가, 32개 정보서비스 분류별로 각종 기술정보 및 다양한 정보서비스를 제공 중에 있다.

〈표 5〉는 지난 2016년부터 OVICE를 이용한 이용자들의 국가별 다운로드/페이지뷰 수 및 데이터의 평균 업로드 기간을 나타낸 표이다. 일반적으로 업로드 기간이 긴 베트남, 인도네시아에 다운로드수가 더 많은 것을 확인 할 수 있지만, 이에 반해 페이지뷰 수는 상대적으로 그 편차가 적은 것을 확인할 수 있다. 따라서 이들을 다시 정보서비스 분류별로 다시 나누어서 분석

을 수행할 경우 충분히 의미 있는 결과를 도출할 수 있을 것으로 보인다.

이를 통해 사용자들의 필요정보 우선순위를 판단하기 위해선 어느 분류에서 높은 이용률을 보였는가를 판단해야 하지만 단순한 다운로드, 페이지뷰 간의 관계만으로는 각종 정보 분류의 명확한 우선순위를 파악하는 데에 어려움이 있다. 이에 각각의 분류에 대한 명확한 우선순위를 매기는 것 보다는 이들의 분포에 따른 비슷한 특성을 보이는 분류를 묶어서 각각의 군집에 대한 특성을 파악하는 것이 더 효과적일 것이다. 이에 본 연구에서는 군집화 분석을 통해 각각의 주제 분류를 페이지뷰 수, 다운로드 수 특성에 따라 분류하고, 이를 통해 해외건설정보수요를 분석하였다.

〈표 5〉 국가별 다운로드/페이지뷰 및 평균 업로드 기간

| 국가 | 다운로드 수 | 페이지뷰 수 | 평균 업로드 기간(개월) |
|--------|--------|--------|---------------|
| 베트남 | 1,177 | 20,906 | 20.09 |
| 인도네시아 | 1,285 | 20,308 | 22.25 |
| 필리핀 | 339 | 8,819 | 15.86 |
| 방글라데시 | 112 | 7,178 | 14.92 |
| 미얀마 | 369 | 11,388 | 16.28 |
| 페루 | 210 | 12,286 | 16.86 |
| 터키 | 224 | 12,323 | 8.35 |
| 케냐 | 159 | 10,848 | 8.53 |
| 미국 | 94 | 13,995 | 7.33 |
| 국가 미분류 | 777 | 33,912 | 17.37 |

2.3 K-Means Clustering Method

군집화(Clustering) 기법이란 분류(Classification)와 구분되는 데이터 분석기법의 한 종류로 데이터 분석에 명확한 기준을 설정하기 힘든 상황에서, 데이터 분포의 특성을 파악하여 유사성을 파악한 뒤 임의의 클래스로 구분하는 방법이다. 그 중 군집화 기법의 가장 대표적인 예가 K-means Clustering이라는 방법이다. 본 방법론은 Euclidean Distance 유사도를 활용하여 데이터들의 물리적 거리를 기준으로 비슷한 거리에 있는 데이터를 k 개의 클래스로 군집화하는 방식이다.

우선 N 개의 데이터를 가지는 데이터 x_n ($n = 1, \dots, N$)에 대하여 K ($k = 1, \dots, K$)개의 군집으로 분류할 때, 주어진 데이터 집합으로부터의 최소가 되는 각 군집의 중심점(u_k)을 구해야 한다. 이를 위해 변수 $r_n^k \in \{0, 1\}$ 을 Eq. (1)과 같이 정의하면,

$$r_n^k = \begin{cases} 1 & \text{if } x_n \in V_k \\ 0 & \text{if } x_n \notin k \end{cases} \quad (1)$$

같은 군집에 속하는 각각의 점들로부터 그 군집의 중심점과의 거리의 제곱의 합(S)은 Eq. (2)와 같다.

$$S = \sum_{n=1}^N \sum_{k=1}^K r_n^k (x_n - u_k)^2 \quad (2)$$

여기서 거리의 합(S)이 최소가 되는 r_n^k 와 u_k 을 구하기 위해서는 다음의 반복적인 과정이 필요하다. 먼저 u_k 의 초기 값을 임의로 설정한다.

다음에 ① 이 u_k 값에 대해 S 가 최소화되는 r_n^k 를 구한다. ② 분류된 r_n^k 간의 무게중심을 구해 새로운 u_k 을 구한다. 앞선 ①, ②의 과정을 계속 반복하여 더 이상 u_k 값이 변하지 않게 될 때의, r_n^k 의 군집이 최적의 군집분류가 된다.

본 방식은 구분하고자 하는 군집 수 k 만을 지정되면 각 군집에 대한 정의를 하지 않아도 최적화된 군집의 분류가 가능해진다. 하지만 각 군집에 대한 정의가 따로 이루어지지 않기 때문에 군집화 이후 각 그룹이 갖는 특성은 사용자가 직접 분석을 통해 정의해야 한다.

K-means clustering에서 또 하나 중요한 부분은 데이터 분포에 적합한 군집 수 k 를 정하는 것이다. 최적의 k 를 구하는 가장 쉬운 방법은 각 군집 수 K 에 대한 분류의 품질을 판단하는 것이다. 이를 판단하기 위한 좋은 방법으로 실루엣 기법이 있다.

실루엣 기법은 군집간의 응집도, 분리도를 측정해 군집의 품질을 정량적으로 판단할 수 있는 방법이다. 만약, 군집간의 분류가 최적으로 이루어 졌다면 클러스터 안의 거리는 짧을 것(응집도)이고, 다른 클러스터와의 거리는 상대적으로 멀 것(분리도)이다. 여기서 데이터 x_n 에 대한 실루엣계수 s_n 은 다음 Eq. (3)과 같이 정의할 수 있다.

$$s_n = \frac{b_n - a_n}{\max(a_n, b_n)} \quad (3)$$

여기서, a_n 은 데이터 x_n 과 동일한 군집내의 나머지 데이터들과의 평균 거리(응집도), b_n 은 x_n 과 가장 가까운 군집내의 데이터들과의 평균 거리이다. 만약 이 s_n 값이 0에 가깝다면 분류가 무의미하다는 의미이고, s_n 값이 1에 가깝다면 최

적의 분류가 이루어 졌다고 판단할 수 있다. 즉, s_n 의 값이 1에 가장 가까운 군집을 구하면 이는 데이터 분포에서 최적의 군집분류가 될 것이다.

보로는 사회문화정치경제, 기술트렌드, 기타일반 등의 순서로 조사되었다

3. 연구 결과

3.1 해외건설정보 필요정보 설문결과

한국건설기술연구원에서는 향후 해외건설 관련 정보시스템이서 구축될 수 있는 필요정보 콘텐츠 개발범위 선정을 위해 해외건설 전문가를 대상으로 우선순위 도출을 위한 설문조사를 실시하였다. 설문조사를 토대로 평가한 정보서비스 분류별 우선순위 <표 6>과 같다.

설문 결과 우선순위가 가장 높은 정보로는 클레임자료, 입찰관련 정보, 업체정보, 발주처정보 등의 순서로 조사되었으며 우선순위가 낮은 정

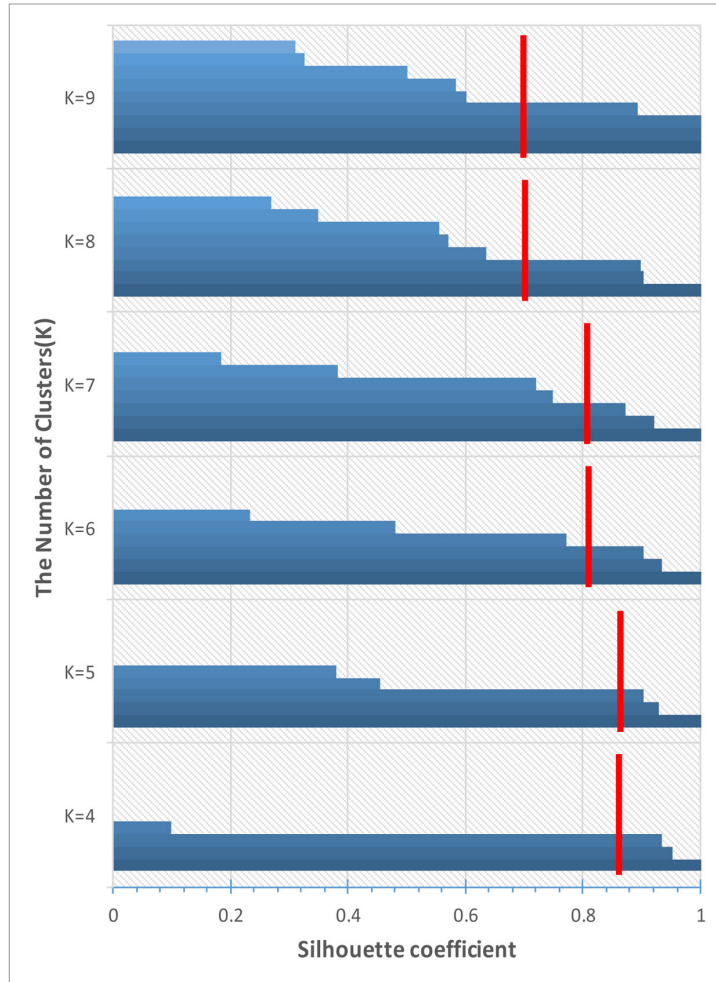
3.2 군집분석을 통한 해외건설정보 우선순위 분류

단순한 다운로드 수, 페이지뷰 수만을 통해 정확한 우선순위를 정하는 것이 어렵기 때문에 군집 분석을 통해 그룹별 특성을 확인하고 그룹별 분포를 통해 기존 설문으로 정했던 우선순위에 대한 평가를 수행하였다. 군집분석을 위해 OVICE 통계를 가지고 앞서 언급했던 K-means Clustering을 이용하였다.

우선 가장 적합한 군집의 개수를 선정하기 위해 통계 분포상 군집으로 엮기 어려운 2, 3개의 군집개수를 제외한 4~9개의 군집으로 분류했을 때의 각 분류별 실루엣계수 분포는 <그림 1>과 같다.

<표 6> 정보서비스 분류별 우선순위 설문결과

| 대분류 | 소분류 | 점수 | 순위 | 대분류 | 소분류 | 점수 | 순위 |
|----------|-----------------|------|---------------|---------|-----------------|------|----|
| 영업 및 기획 | 시장동향 | 3.67 | 21 | 조사 및 설계 | 기본설계/계획 도서정보 | 4.00 | 14 |
| | 사회문화정치경제 | 2.90 | 32 | | 인허가정보 | 3.95 | 15 |
| | 일반 법령 및 제도 | 3.95 | 15 | | 프로젝트별 실시설계 도서정보 | 3.90 | 18 |
| | 건설 법령 및 제도 | 4.19 | 9 | 사업관리 | 구매조달 | 3.90 | 18 |
| | 사업개발계획 | 4.33 | 6 | | 일반관리 | 3.67 | 21 |
| | 건설인프라 현황 및 환경정보 | 3.57 | 26 | | 발주처관리 | 4.43 | 4 |
| | 건설시장동향 | 3.57 | 26 | | 공정 및 원가관리 | 4.05 | 13 |
| | 기술트렌드 | 3.05 | 31 | | 품질안전보건환경 | 3.62 | 24 |
| 입찰 및 계약 | 입찰관련 정보 | 4.71 | 2 | | 협력업체관리 | 3.81 | 20 |
| | 입찰관련 문서 | 4.38 | 5 | | 운영 및 유지관리 | 3.62 | 24 |
| | 보증 및 보험 | 4.10 | 11 | 공통일반 | 분쟁과 협의 및 클레임자료 | 4.76 | 1 |
| | 현장조사 | 4.14 | 10 | | 업체정보 | 4.52 | 3 |
| | 견적 및 물가자료 | 4.29 | 7 | | 코드 및 기준 | 3.67 | 21 |
| 계약 관리 정보 | 4.24 | 8 | 지역, 국가별 리스크현황 | | 3.95 | 15 | |
| 조사 및 설계 | 지형/지질/기후 | 3.29 | 30 | | 성공 및 실패사례 | 4.10 | 11 |
| | 조사보고서 | 3.43 | 28 | 기타일반 | 3.33 | 29 | |



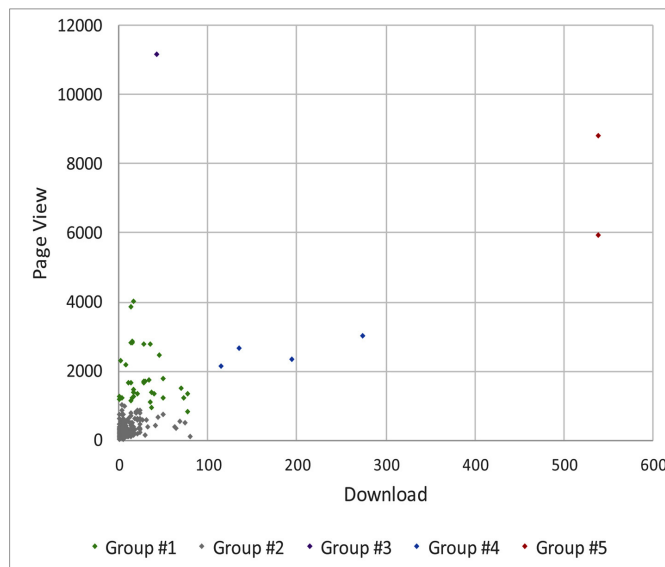
〈그림 1〉 각 군집별 실루엣 계수

〈그림 1〉에서 각 항목별 평균 실루엣계수(실선)를 살펴보면 군집 개수가 커질수록 군집 간 응집도가 큰 군집의 개수가 증가하지만 그만큼 분리도가 커지는 군집의 개수도 같이 증가함을 알 수 있다. 이를 수치상으로 비교할 때, 〈표 7〉과 같이 군집의 개수가 5개일 때 최적의 분류가 이루어졌음을 알 수 있다. 또한, 최적의 군집수인 5개의 군집으로 분류한 320개 항목별 다운로드/페이지뷰 분포는 〈그림 2〉와 같다.

그래프 상에 좌측 하단에 분포된 Group #1, Group #2는 적은 다운로드/페이지뷰 수 분포를 보이고 있어 이 두 군집에 분포한 분류항목들은 우선순위를 고려할 때 큰 의미를 가지기 어렵다고 판단할 수 있다. 반면에 상대적으로 높은 값을 가지는 Group #3~5의 분류항목들은 상대적으로 의미를 가지고 있다고 판단할 수 있다. 〈표 8〉은 Group #3~5에 해당하는 분류항목을 나타낸 표이다.

〈표 7〉 군집개수별 평균 실루엣 계수

| 군집개수 | | K=4 | K=5 | K=6 | K=7 | K=8 | K=9 |
|------------------------|----|--------|--------|--------|--------|--------|--------|
| 군집별 평균 실루엣 계수 | #1 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | #2 | 0.9510 | 0.9299 | 0.9345 | 0.9209 | 0.9027 | 1.0000 |
| | #3 | 0.9334 | 0.9027 | 0.9027 | 0.8735 | 0.8990 | 1.0000 |
| | #4 | 0.0984 | 0.4560 | 0.7718 | 0.7493 | 0.6357 | 0.8929 |
| | #5 | | 0.3790 | 0.4795 | 0.7214 | 0.5717 | 0.6011 |
| | #6 | | | 0.2321 | 0.3830 | 0.5558 | 0.5835 |
| | #7 | | | | 0.1850 | 0.3491 | 0.5019 |
| | #8 | | | | | 0.2690 | 0.3262 |
| | #9 | | | | | | 0.3093 |
| 항목별 실루엣 계수 평균 | | 0.8631 | 0.8638 | 0.8096 | 0.8077 | 0.7016 | 0.7007 |



〈그림 2〉 다운로드/페이지뷰 분포에 따른 군집분류 결과(k=5)

〈표 8〉 Group #3-5에 해당하는 분류항목

| Cluster | 국가분류 | 주제분류 | 다운로드 수 | 페이지뷰 수 |
|----------|-------|--------------|--------|--------|
| Group #3 | - | 입찰관련문서 | 42 | 11,135 |
| Group #4 | 베트남 | 일반관리 | 115 | 2,121 |
| Group #4 | 베트남 | 건설법령 및 제도 | 194 | 2,325 |
| Group #4 | 인도네시아 | 조사보고서 | 135 | 2,633 |
| Group #4 | 베트남 | 조사보고서 | 274 | 2,996 |
| Group #5 | 인도네시아 | 기본설계/계획 도서정보 | 538 | 5,924 |
| Group #5 | - | 기타일반 | 538 | 8,792 |

Group #3은 다른 자료들과 비교하여 좌상 방향으로 편향된 모습을 보이고 있다. 이는 필요에 의해 해당 페이지에 접속한 횟수는 많지만 상대적으로 자료를 다운로드하여 사용한 횟수는 적다는 것을 의미한다. 해당 분류에 속하는 자료는 국가에 상관없는 다자간개발은행(WB, ADB 등)의 입찰관련 서류 또는 일반적으로 사용되는 입찰관련문서를 제공하고 있다. 이는 앞선 정보서비스 분류별 우선순위 설문결과에서도 필요우선순위가 높게 나타난 것으로 실제 사용자들의 수요만큼 필요한 서비스 제공이 이루어지지 않고 있다고 판단할 수 있다.

Group #4는 다른 자료들과 비교하여 상대적으로 다운로드 수가 많은 자료로, Group #3과 비교했을 때 접속 수에 비해 다운로드가 많이 이루어진 항목으로 볼 수 있다. 이들 항목은 우선순위 설문결과에서는 필요 우선순위가 높지 않았던 항목이었음에도 불구하고 다운로드 수치가 높게 나타난 것으로 정보제공의 품질이 상대적으로 높았다고 볼 수 있다. 특히 해당 군집의 국가는 자료 기간이 긴 베트남, 인도네시아가 대부분이었는데, 이를 통해 일반관리, 건설법령 및 제도, 조사보고서 등의 자료는 실제 사용자들이 오랜 기간 꾸준히 이용한 자료라고 판단할 수 있다. 따라서 추후 정보시스템의 정보제공의 관점에서 이들 주제에 대한 지속적인 정보업로드는 꾸준한 이용자의 접속을 유도할 수 있을 것이다.

Group #5는 가장 높은 다운로드를 기록한 군집이다. 우선 인도네시아 기본설계/계획 도서정보의 경우 우선순위 필요정보 설문 결과가 높지 않으나, 해당 항목의 특징을 살펴보면 다른 국가의 기본설계/계획 도서정보보다 자료의

양이 매우 많으며 정보제공도 체계적으로 잘 정리되어 있다. 이는 필요 정보의 우선순위가 높지 않더라도 제공 품질이 좋은 자료들에 대해서는 많은 사용자가 이용하고 있다는 것으로 해석할 수 있어, 향후 정보서비스 운영에 있어서 제공 정보의 품질향상 또한 매우 중요하다고 볼 수 있다. 국가 미분류의 기타일반 자료의 경우 각종 사업관리, 영문레터 등의 서식/양식 자료가 많이 제공되고 있다. 필요정보 우선순위 설문 과정에서 기타일반이라는 이름으로 인해 실제 우선순위는 높게 나타나지 않았으나, 실무과정에서 필요한 각종 양식/서식의 제공은 실제 사용자들의 수요가 상당히 많다는 것으로 판단할 수 있다.

4. 결 론

본 연구에서는 분류체계별 필요정보 우선순위를 선정함에 있어 OVICE 이용통계 및 해외 건설정보 수요를 비교·분석 하여 정보시스템의 정보제공 방향성에 대한 판단을 위한 기초자료로 활용하고자 하였다. 또한 군집화 분석을 통해 각각의 주제 분류를 페이지뷰 수, 다운로드 수 특성에 따라 분류하여 이용자 통계분석의 효율성을 높이고자 하였다. 그 결과, 실제 설문과 사용자 이용통계에는 많은 차이점을 보였는데, 이는 실제 사용자들이 필요로 하는 정보와 사용자가 이용하는 정보가 다르다는 것으로 두 가지 방향으로 해석이 가능하다. 첫째는 정보시스템이 충분히 사용자를 유인할 만한 정보제공이 이루어지지 않았다고 볼 수 있으며, 다른 관점으로는 일부 전문가를 대상으로 하는 설문 결

과가 전체사용자의 수요를 대변하지 못했다고 해석할 수 있다. 하지만 수요조사 결과와 실제 이용통계 우선순위가 유일하게 일치한 입찰관련문서 항목에서 상대적으로 페이지 뷰에 비해 다운로드수가 매우 적었던 것을 보면 실제 사용자들이 필요로 하는 정보에 맞춰 충분한 서비스가 이루어지지 않고 있다고 해석하는 것이 정확하다고 볼 수 있다.

따라서 향후 정보서비스 운영과정에서는 앞서 언급된 필요 우선순위가 높은 항목들에 대한

정보서비스 품질을 높일 필요가 있으며, 이와 더불어 사용자 통계에서 높은 수치를 보여준 항목들을 참고하여 서비스 제공의 품질을 높이게 된다면 국내 건설엔지니어링 기업들의 해외건설 경쟁력 강화라는 목표를 이루는데 큰 도움이 될 것이다. 더불어, 향후 더 많은 관련 자료의 축적과 함께 이용자의 접속이 꾸준히 이루어지게 된다면 추후 실제 사용자들의 수요분석에 좀 더 정확한 평가가 가능할 것이며 더 좋은 품질의 정보서비스 제공이 가능할 것으로 사료된다.

참 고 문 헌

- 권오철, 조정근. 2017. 해외건설엔지니어링시스템 개발을 위한 정보수요조사 연구. 『대한건축학회 논문집-구조계』, 33(4): 37-46.
- 박환표, 신은영, 이교선. 2008. 국내 건설업체의 해외진출 활성화 방안. 『한국건설관리학회논문집』, 9(6): 225-234.
- 백승원, 한승헌, 이창준, 이지섭, 문수환. 2018. 비정형데이터 기반 공공 건설사업 갈등 이슈 분석-제주해군기지 사례를 중심으로. 『공공사회연구』, 8(1): 83-106.
- 윤승희, 유정호, 조훈희, 장현승. 2015. 해외건설 동반진출을 위한 중소건설업체의 역량확보 방안. 『대한건축학회 논문집-구조계』, 31(5): 97-104.
- 이지섭, 김도윤, 이창준, 이정훈, 한승헌. 2018. 군집분석을 통한 공공 건설사업 갈등 유형화 연구. 『한국건설관리학회 논문집』, 19(2): 61-72.
- 한국건설기술연구원. 2017. 『해외건설엔지니어링 정보시스템 중장기 운영전략계획 수립』. 고양: 한국건설기술연구원.
- Ketchen, D. J. and C. L. Shook. 1996. "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique." *Strategic Management Journal*, 17(6): 441-458.
- MacQueen, James. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14): 281-297.

- Martin, N. and H. Maes. 1979. *Multivariate Analysis*. Massachusetts: Academic press.
- Rousseeuw, Peter J. 1987. "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics*, 20: 53-65.

• 국문 참고자료의 영어 표기

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

- Baek, Seung Won, Seung Heon Han, Changjun Lee, Ji Seop Lee, and Soo Hwan Moon. 2018. "Conflict Analysis in Public Construction Project with Unstructured Data: The Case of Jeju Naval Base Project." *Journal of Public Society*, 8(1): 83-106.
- KOREA Institute of Civil Engineering and Building Technology. 2017. *Long-term Information Strategy Planning of Overseas Information System for Construction Engineering*. Goyang: KOREA Institute of Civil Engineering and Building Technology.
- Kwon, O-Cheol and Jeong-Keun Cho. 2017. "Research on Information Demand for Development of Overseas Construction Engineering System." *Journal of the Architectural Institute of Korea*, 33(4): 37-46.
- Lee, Jiseop, Doyun Kim, Changjun Lee, Jeonghun Lee, and Seungheon Han. 2018. "A Research for Clustering of Conflict in Public Construction Project." *Korean Journal of Construction Engineering and Management*, 19(2): 61-72.
- Park, Hwan-Pyo, Eun-Young Shin, and Kyo-Sun Lee. 2008. "Improvement plan going into other countries of domestic construction companies." *Korean Journal of Construction Engineering and Management*, 9(6): 225-234.
- Yun, S., J. Yu, H. Cho, and H. Jang. 2015. "A Study on Strategies of Small and Midsize Construction Companies for Joint Overseas Expansion." *Journal of the Architectural Institute of Korea*, 31(5): 97-104.