

소셜 빅데이터분석을 통한 외국인근로자에 관한 국민 인식 분석과 정책적 함의

하재빈¹, 이도은^{2*}

¹아주대학교 비즈니스애널리틱스전공 박사수료, ²고려대학교 법학연구원 연구원

Analysis of Public Perception and Policy Implications of Foreign Workers through Social Big Data analysis

Jae-Been Ha¹, Do-Eun Lee^{2*}

¹Complete a doctorate, Business Analytics and Information Systems, Ajou University

²Researcher, Korea University Legal Research Institute

요 약 본 연구에서는 빅데이터 기법 중에 하나인 텍스트마이닝을 활용하여 소셜플랫폼에서 외국인근로자에 대한 인식을 알아보고 시사점을 도출하고자 하였다. 연구를 위해서 2020년 1월 1일부터 12월 31일까지를 기준으로 '외국인근로자' 검색 키워드를 수집하여 빈도분석, TF-IDF 분석, 연결중심성 분석으로 상위 키워드 100개를 도출하고 비교분석을 수행하였다. 또한 Ucinet6.0과 Netdraw를 이용해 의미연결망을 분석하였으며, CONCOR 분석을 통해 외국인정책 이슈, 지역사회 이슈, 사업주 관점 이슈, 고용 이슈, 근로환경 이슈, 법적 이슈, 출입국 이슈, 인권 이슈로 8개 클로스터로 군집화하였다. 이러한 분석 결과를 바탕으로 외국인근로자 국민적 인식, 주요 이슈를 파악하였으며, 향후 외국인근로자에 대한 정책 및 관련 연구에 필요한 기초자료를 제공하고자 한다.

주제어 : 외국인근로자, 소셜 빅데이터분석, 텍스트마이닝, TF-IDF, 연결중심성, 의미연결망분석

Abstract This paper aimed to look at the awareness of foreign workers in social platforms by using text mining, one of the big data techniques and draw suggestions for foreign workers. To achieve this purpose, data collection was conducted with search keyword 'Foreign Worker' from Jan. 1, to Dec. 31, 2020, and frequency analysis, TF-IDF analysis, and degree centrality analysis and 100 parent keywords were drawn for comparison. Furthermore, Ucinet6.0 and Netdraw were used to analyze semantic networks, and through CONCOR analysis, data were clustered into the following eight groups: foreigner policy issue, regional community issue, business owner's perspective issue, employment issue, working environment issue, legal issue, immigration issue, and human rights issue. Based on such analyzed results, it identified national awareness of foreign workers and main issues and provided the basic data on policy proposals for foreign workers and related researches.

Key Words : Foreign Worker, Social Big Data Analysis, Text Mining, TF-IDF, Degree Centrality, Semantic Network Analysis

*Corresponding Author : Do-Eun Lee(favorite6844@hanmail.net)

Received August 11, 2021

Accepted November 20, 2021

Revised October 20, 2021

Published November 28, 2021

1. 서론

우리나라 국민이 가지고 있는 외국인에 대한 인식은 상당 부분 과도기적인 양상을 띄고 있지만, 시간이 지날수록 국민들이 외국인근로자를 접할 기회가 점차 많아지면서 외국인근로자에 대한 친밀도가 높아졌다[1]. 다만 세대별로 외국인근로자에 국민적 인식은 다양하게 존재하고 있다. 예를 들어 20대 국민의 경우 외국인근로자 유입으로 인한 장점인 문화적 다양성을 인지하고 있으나, 동시에 외국인근로자를 타자화하여 한국문화와 동화되어야 하는 대상으로 인식하고 있다. 반면에 50대의 경우 외국인근로자의 경제적 당위성과 불가피성을 인지하고 있으며, 이를 객관화함으로써 외국인근로자에 대한 부정적 인식에 국민의 편견이나 오류가 존재함을 인정하고 있다. 20대와 50대 모두 자국민 취업을 위한 근로조건 개선, 외국인근로자에 대한 강력한 선별 및 관리 정책의 필요성에 관해서는 같은 의견을 내고 있다[2]. 언론보도에 있어서도 외국인근로자에 대한 보도성향은 부정적인 기사의 비율이 다소 높은 것으로 나타났다[3].

외국인근로자에 대한 다양한 연구가 수행되고 있지만 소셜네트워크서비스 상에서 발생하는 빅데이터를 기반으로 한 인식 연구는 미흡한 실정이다. 소셜네트워크서비스 빅데이터 분석을 정책 제언 과정에 활용할 경우 국민 공감형 정책이나 미래 변화에 대응한 정책 개발이 가능하기 때문에[4], 본 연구에서 외국인근로자에 관한 소셜 빅데이터를 분석하여 정책입안자 또는 연구기관이 활용할 수 있는 외국인근로자 인식 관련 기초자료를 제공하고자 한다.

2. 선행연구

2.1 외국인근로자

모든 이주노동자와 그 가족들의 보호를 위한 국제협약(The International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families) 제2조에서 외국인 근로자와 일맥상통하는 이주노동자에 대해 정의하고 있다. 이주노동자란 “국적국이 아닌 나라에서 유급활동에 종사할 예정이거나 이에 종사하고 있거나 또는 종사하여 온 사람”이다. 국내법인 외국인 고용 등에 관한 법률 제 2조에 의하면 외국인근로자라 함은 “대한민국의 국적을 가지지 아니한 사람으로서 국내에 소재하고 있는 사업 또는 사업장에서 임금을 목적으로 근로를 제공하고 있거나 제공하려는 사람”을 의미한다. 본 연구에서는 국내법에서 규정하고 있는 법적 개념인 ‘외국인근로자’를 분석 키워드로 선정하였다.

한편 한국에서 외국인근로자는 생명권, 고문의 금지와 같은 자유권을 향유할 수 있으나 근로의 권리, 사회보장에 대한 권리 등을 포함하는 사회권적 기본권은 체류자격에 따라 제한될 수 있다[5]. 외국인을 헌법상 기본권의 주체로 볼 수 있지만 공익을 이유로 외국인의 직업선택의 자유 및 영업의 자유 등은 무제한적으로 누릴 수 없는 것이다[6]. 아래 Fig 1은 2017년부터 2019년까지 3년간 체류자격별 등록외국인의 수를 나타내고 있다[7,8,9]. 비전문취업(E-9)으로 등록된 외국인이 가장 높은 비율을 차지하며, 다음으로 방문취업(H-2)으로 등록된 외국인이 두 번째로 높은 비율을 차지하였다.

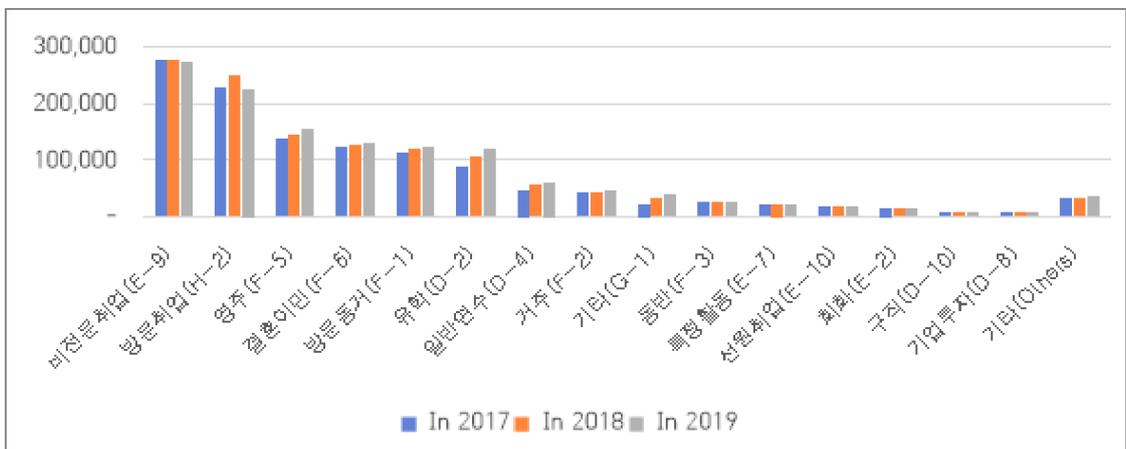


Fig. 1. Status of Registered Foreigners by Status of Stay (2017-2019)

2.2 텍스트마이닝

비정형 데이터의 처리와 분석, 인사이트 도출, 의사결정 자료 제공 등의 기법으로 각광받고 있는 기법이 바로 텍스트마이닝(Text Mining)이다[10]. 텍스트마이닝은 1995년 텍스트 기반 지식 발견(knowledge discovery in textual database)의 개념으로 최초 언급되었다[11]. 즉 텍스트마이닝 기법이란 비정형의 데이터 형식인 텍스트를 기반으로 한 자연어 처리 기술을 통해 유용한 패턴과 정보를 추출하고 새로운 지식을 발견하는 것을 의미한다[12].

빈도분석(frequency analysis)은 분류하고자 하는 자료에서 특정 단어가 얼마나 자주 등장하는지를 지표로 나타낸 방법이므로 단어의 빈도가 높을수록 해당 단어가 핵심적인 단어임을 의미하게 된다[13]. 빈도분석을 시각화하는 기법인 워드 클라우드(Word cloud)는 키워드의 빈도에 따라서 문자의 크기를 결정하여 텍스트의 빈도를 직관적으로 이해할 수 있도록 하는 특징을 가지고 있다[14].

TF-IDF(Term Frequency - Inverse Document Frequency)의 경우 정보의 검색과 머신러닝 등에 많이 활용되고 있다. 단어 중요도의 가중치를 바탕으로 여러 문서로 이루어진 문서 군에서 특정 단어가 하나의 문서에서 얼마나 중요한 영향을 미치는지를 보여주는 통계적인 수치를 의미한다. 한 논문에서 단지 한 단어에 가중치를 두어 다른 정교한 텍스트 표현보다 텍스트 색인에 있어서 우수한 결과를 얻을 수 있다[15]. TF(Term Frequency, 단어빈도)는 특정한 단어가 문서 내에서 얼마나 자주 등장하는지를 나타내는 값인 DF(문서 빈도, document frequency)와 이 값의 역수인 IDF(역문서 빈도, inverse document frequency)로 나뉜다. 예를 들어서 ‘외국인근로자’라는 단어가 있을 때 ‘외국인’이나 ‘근로자’와 같은 단어가 거의 모든 문서에서 존재하게 되지만 단순한 정보로서의 가치는 떨어지게 된다. 이를 보완하기 위해 나온 개념이 바로 IDF이다. 특정한 문서에서 단어의 빈도가 높지만 전체 문서에 해당 단어를 포함하는 문서가 적을수록 TF-IDF의 값이 높아진다. 즉, TF-IDF의 값이 높을수록 중요한 텍스트가 되어 여러 문서의 집합으로부터 특정 단어가 특정 문서에서 얼마나 중요한가를 판단하는 기준이 된다[16]. TF-IDF의 수식은 다음과 같다.

$$TF-IDF = TF \times \frac{1}{IDF}$$

TF: 문서 내 특정 단어 빈도수

DF: 여러 문서 내 특정 단어 빈도수

IDF: DF의 역수

2.3 의미연결망 분석

의미연결망 분석(Semantic Network Analysis)은 커뮤니케이션 내용에서 단어가 동시에 발생하는 빈도를 측정하여 단어 간의 의미 해석과 상징적인 연결구조를 파악할 수 있게 한다[17,18,19]. 실제로 의미연결망 분석은 사회적인 시스템 구조를 파악하고자 할 때 행위자 간의 관계를 중시하는 사회적연결망 분석 기법을 커뮤니케이션 메시지에 적용한 방법이다[20]. 본 연구에서는 사회적연결망에서 가장 대표적인 개념인 연결 중심성(Degree Centrality) 및 CONCOR(convergence of iterated correlations) 분석을 실시하였다.

중심성의 경우 연결망 내에서 특정 개체가 중심에 위치하는 정도를 나타낸다(Freeman, 2008). 연결 중심성 분석(Degree centrality Analysis)은 여러 연결 관계를 양적으로 분석하여 영향력을 가능하는 방법이다. 많은 연결 관계를 가질수록 연결 중심성이 높아진다는 특징을 가지고 있다[21].

CONCOR(Convergence of iterated correlations) 분석은 구조적 동위성(Structural Equivalence)분석 방법의 하나며, 상관관계 분석을 반복적으로 실행하여 적정 수준의 집단을 찾아내는 방법으로, 반복적으로 생성되는 상관계수(Iterated correlations) 매트릭스의 셀의 값이 ‘+1’ 혹은 ‘-1’이 수렴(Convergence)하는 순간까지 지속해서 반복되며, 최종 매트릭스의 값이 ‘+1’과 ‘-1’만의 값을 가질 때까지 지속해서 반복된다[22]. 과거의 텍스트 빈도 분석의 경우 연구자의 주관이 배제되기 어려웠지만[23], 의미연결망분석을 통하게 되면 미시적 관점에서 단어 간의 관계 패턴으로 구성된 구조를 파악할 수 있게 되기 때문에 본 연구에서는 의미연결망분석을 바탕으로 외국인근로자와 관련한 단어와 단어 간의 관계와 위치를 확인해보고자 한다.

3. 연구내용 및 연구방법

3.1 연구문제

외국인근로자에 대한 국민적인 인식과 트렌드를 분석하기 위해 2020년 1월부터 2020년 12월까지 1년 동안 국내 주요 포털인 네이버와 다음의 콘텐츠에서 ‘외국인근로자’와 관련된 키워드를 수집하였다. 국내 포털 사용자들을 중심으로 많이 언급된 키워드를 이용해 이슈와 인식을 파악하고 키워드 간의 연결성의 특징을 분석하고자 한다.

- 연구문제 1: 외국인근로자 관련 소셜 빅데이터에 나타난 주요한 단어는 무엇인가?
- 연구문제 2: 외국인근로자 관련 소셜 빅데이터에 나타난 주요한 단어 연결망은 어떠한 구조적 특성을 보이는가?
- 연구문제 3: 외국인근로자 관련 단어 연결망의 특성은 어떠한 의미를 구성하는고 있는가?

3.2 연구대상

외국인근로자의 텍스트마이닝 분석을 실시하기 위해 네이버(NAVER)와 다음(DAUM)의 뉴스, 블로그, 카페, 지식인의 검색데이터를 분석데이터로 선정하였다. 국내에서 가장 인지도가 높은 포털인 네이버와 다음을 분석 대상으로 하여, 소셜 빅데이터 분석을 통한 인식 연구를 수행한 선행연구[24-26]를 참고하였으며, 본 연구에서는 국민을 대상으로 외국인근로자에 관한 인식을 연구하는 것을 목적으로 하고 있으므로 해외 사용자가 많은 해외 유명 소셜네트워크서비스인 구글, 유튜브, 트위터와 같이 글로벌 플랫폼은 분석 대상에서 제외하였다. 분석키워드는 ‘외국인근로자’로 설정하였으며, 분석 기간은 2020년 1월 1일부터 12월 31일까지로 총 1년으로 설정하였다. 보다 객관적인 분석을 위해 법학전공 전공자 1인과 공동으로 단어 정제를 수행하여 신뢰성을 높이고자 하였으며, 상위 100개의 단어를 추출해 네트워크의 가시성을 확보하였다.

3.3 분석방법

본 연구를 위해 텍스트(TEXTOM) 솔루션을 사용하여 텍스트마이닝 분석을 실시하였다. 텍스트는 소셜데이터의 수집과 정제, TF-IDF 분석, 연결중심성 분석, 시각화 등 데이터의 일괄 처리를 제공하는 솔루션이다. 또한, Ucinet6.0 프로그램을 이용하여 단어 간의 연결 관계와 연결 강도, 전체 연결망에 차지하는 단어의 위치를 파악할 수 있는 중심성 지수를 분석하였으며, NetDraw를 바탕으로 CONCOR 분석을 수행하여 시각화 작업을 완료하였다.

4. 연구결과

4.1 데이터 수집 결과

2020년 1월 1일부터 2020년 12월 31일까지 ‘외국인

근로자’ 키워드를 중심으로 네이버, 다음에서 텍스트를 추출하였다. 구체적으로 네이버 3,625건(블로그 999건, 카페 999건, 지식인 627건, 뉴스 1,000건), 다음 2,592건(블로그 889건, 카페 984건, 뉴스 719건)의 데이터가 수집되었으며, 총 1,453,272개 단어가 모였다. 아래 Table 1은 데이터 정제의 예시이다.

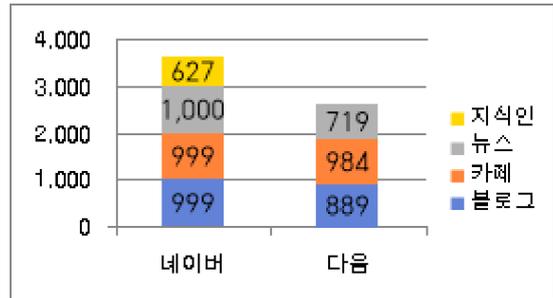


Fig. 2. Data Collection

4.2 데이터 정제작업

데이터 분석의 신뢰성을 향상시키기 위해 데이터 정제 작업을 수행하였다. 구체적으로 외국인근로자에 관하여 수집된 단어 중 무관한 단어는 삭제하였다. 검색 키워드와 동일한 ‘외국인근로자’ 키워드 역시 제거하였다. 또한, 띄어쓰기가 잘못되거나 불완전한 상태인 단어는 수정하였으며, 의미가 같은 단어들은 통합하는 정제작업을 시행하였다. 아래 Table 1은 데이터 정제의 예시이다.

Table 1. Data Cleansing

Classification	Data Cleansing
Delete unnecessary words	2, 3, ABC etc.
correct spacing	Work place-> workplace etc.
Combine words with the same meaning	Seoul Metropolitan Government -> Seoul Cheonan-si -> Cheonan Incheon Metropolitan City -> Incheon etc.

4.3 데이터 분석 결과

4.3.1 데이터 분석 결과

외국인근로자 키워드의 빈도분석 결과 빈도가 높은 상위 100개의 키워드는 다음 Table 2과 같다. 외국인, 일 자리, 기업, 사용, 근로자, 코로나19, 서울특별시, 사회, 서비스, 이용, 정부, 법률, 출입국, 지역, 문화, 교육, 사업, 국민, 고용, 사업주 등의 순으로 빈도가 높게 나타났다.

Table 2. Frequency Analysis of Foreign Workers

WORD	Frequency Analysis	WORD	Frequency Analysis	WORD	Frequency Analysis	WORD	Frequency Analysis
외국인	8,397	정책	1,657	건강	1,211	사건	903
일자리	7,813	여성	1,638	대통령	1,202	개선	894
기업	6,237	사업장	1,634	인천광역시	1,195	강화	893
사용	5,484	경기	1,599	환경	1,179	부산광역시	892
근로자	5,238	중소기업	1,570	국회	1,165	남성	891
코로나19	4,548	중국	1,568	대전광역시	1,162	건설	890
서울특별시	4,125	취업	1,496	안전	1,161	무단	889
사회	4,042	감염	1,456	확산	1,160	금융	888
서비스	3,444	신고	1,454	거주	1,148	청년	888
이용	3,215	시각	1,449	입금	1,120	근로	887
정부	3,197	백신	1,445	대구광역시	1,101	사랑	887
법률	3,097	상담	1,438	발급	1,100	소득	886
출입국	2,830	규정	1,435	기술	1,085	노동	876
지역	2,789	현장	1,424	화재	1,030	자녀	871
문화	2,631	충청북도	1,419	일본	1,023	인사	866
교육	2,546	보호	1,405	공무원	1,023	기관	864
사업	2,412	노동자	1,379	베트남	1,015	전문가	855
국민	2,407	직원	1,347	개발	1,001	고용노동부	853
고용	2,106	비자	1,307	프로그램	1,001	생산	852
사업주	1,886	법률개정	1,288	마스크	987	계약	832
태국	1,880	근무	1,262	업무	984	공급	828
공감	1,866	가족	1,255	인권	941	교회	823
확진	1,765	충청남도	1,239	세종	940	시설	823
천안	1,692	취득	1,224	법률위반	936	사용자	820
산업	1,692	조사	1,212	사망	934	농축산업	816

Fig. 3은 빈도분석을 바탕으로 관련 키워드를 시각화한 워드클라우드이다.



Fig. 3. Foreign Workers Word Cloud

Table 3은 외국인근로자와 관련하여 수집된 문서에서 해당 키워드가 얼마나 중요인지 보여주는 TF-IDF 분석을 통해 추출된 상위 100개의 키워드이다. 기업, 일자리, 외국인, 근로자, 코로나19, 서비스, 사회, 사용, 서울

특별시, 출입국, 정부, 법률, 태국, 문화, 교육, 지역, 사업, 천안, 국민, 이용 등의 순으로 TF-IDF가 높게 나타났다.

외국인근로자 데이터에서 네트워크상 키워드의 영향력을 파악할 수 있는 연결 중심성 분석을 수행한 결과는 다음 Table 4와 같다. 일자리, 외국인, 기업, 근로자, 코로나19, 정부, 서울특별시, 사용, 법률, 지역, 국민, 사회, 사업, 이용, 사업주, 교육, 출입국, 정책, 고용, 서비스 순으로 연결중심성이 높게 나타났으며, 이 키워드들은 다른 키워드와 연결성이 높고, 이에 따라 의무구조와 맥락을 형성하여 의미를 확장시키는 주요한 역할을 한 것으로 드러났다.

외국인근로자 키워드에 대한 빈도분석, TF-IDF 분석, 연결 중심성 분석을 통해 각각 도출된 상위 100개의 키워드 중 중복된 키워드는 총 74개이며, 기업, 근로자, 사용, 서울특별시, 사회, 법률, 서비스, 교육, 사업, 국민, 사업주, 고용, 문화, 사업장, 규정, 노동자, 산업, 상담, 법률개정, 가족 등이 세 가지 분석에서 평균적으로 상위 키워드에 오른 것으로 나타났다.

Table 3. TF-IDF Analysis of Foreign Workers

WORD	TF-IDF	WORD	TF-IDF	WORD	TF-IDF	WORD	TF-IDF
기업	11,869.97	사업장	4,526.47	국회	3,417.22	자격증	2,989.16
일자리	10,841.00	여성	4,525.57	가족	3,370.64	법률위반	2,975.84
외국인	10,578.54	중국	4,432.42	교회	3,363.89	사랑	2,932.22
근로자	8,557.62	충청북도	4,431.44	발급	3,318.18	업무	2,925.32
코로나19	8,227.50	노동자	4,426.83	마스크	3,300.64	공무원	2,920.01
서비스	8,095.96	신고	4,322.58	베트남	3,299.64	사망	2,911.95
사회	7,993.29	산업	4,317.67	조사	3,292.51	청년	2,881.53
사용	7,804.70	공감	4,300.70	인천광역시	3,260.97	근로	2,878.28
서울특별시	7,497.16	백신	4,279.68	보호	3,252.85	세종	2,876.11
출입국	7,225.82	규정	4,272.56	근무	3,246.72	한국어	2,869.31
정부	6,941.46	정책	4,270.44	대구광역시	3,205.97	파산	2,851.08
법률	6,855.05	법률개정	4,249.78	거주	3,205.72	다문화	2,842.71
태국	6,342.88	시장	4,240.90	기술	3,201.47	소득	2,841.70
문화	6,210.32	비자	4,148.17	일본	3,189.43	계절근로	2,836.08
교육	5,924.30	취득	4,085.25	인권	3,184.33	친구	2,797.81
지역	5,752.20	취업	3,885.22	확산	3,183.98	사용자	2,772.10
사업	5,699.26	경기	3,870.67	안전	3,180.51	노동	2,768.31
천안	5,536.01	상당	3,784.26	건강	3,156.32	자유	2,748.26
국민	5,412.78	대통령	3,671.88	남성	3,148.46	자녀	2,747.80
이용	5,389.85	충청남도	3,664.07	이주노동자	3,114.87	징역	2,741.65
확진	5,109.84	회계	3,631.56	사건	3,102.86	생산	2,732.65
고용	5,065.52	현장	3,615.01	환경	3,095.16	제주특별자치도	2,718.77
사업주	4,760.60	직원	3,523.44	개발	3,026.61	방곡	2,703.51
중소기업	4,660.40	임금	3,503.54	프로그램	3,014.85	이슬람	2,683.36
감염	4,613.09	대전광역시	3,441.51	노동조합	3,006.12	금융	2,668.31

Table 4. Degree Centrality of Foreign Workers

WORD	Degree Centrality	WORD	Degree Centrality	WORD	Degree Centrality	WORD	Degree Centrality
일자리	0.031962	여성	0.006359	집	0.00481	사고	0.00385
외국인	0.022696	중소기업	0.006275	전문가	0.004804	국회	0.003844
기업	0.018644	중국	0.006252	공무원	0.004754	사망	0.003844
근로자	0.018268	가족	0.006213	사건	0.004742	학생	0.003839
코로나19	0.015299	사업장	0.006084	임금	0.004726	자녀	0.003822
정부	0.015069	확진	0.005915	개신	0.004703	작업	0.003805
서울특별시	0.012162	신고	0.00591	태국	0.004703	사랑	0.003805
사용	0.012061	조사	0.005814	베트남	0.004675	비용	0.003783
법률	0.011909	직원	0.005797	산업	0.004653	시설	0.003766
지역	0.011775	대통령	0.005792	기관	0.00463	채용	0.003766
국민	0.011342	상당	0.005685	고용노동부	0.004613	파산	0.00376
사회	0.011309	취업	0.005668	공급	0.004512	백신	0.003749
사업	0.011157	법률개정	0.005663	소득	0.004467	청년	0.003738
이용	0.011022	취득	0.00564	자격	0.004445	제조업	0.003715
사업주	0.009558	업무	0.005427	서류	0.0044	노동조합	0.00371
교육	0.008228	일본	0.005298	환경	0.00431	경기	0.003704
출입국	0.008093	개발	0.005276	사업체	0.004299	교회	0.003704
정책	0.007824	안전	0.005175	건강	0.004215	학교	0.003704
고용	0.007818	프로그램	0.005169	생산	0.004007	노동	0.003682
서비스	0.007784	발급	0.005135	강화	0.003996	감염	0.003665
규정	0.006892	비자	0.005096	인권	0.003985	자유	0.003637
시장	0.006869	문화	0.005079	보호	0.003979	대학	0.003631
노동자	0.006606	거주	0.005001	피해	0.003906	농축산업	0.003614
현장	0.006415	기술	0.004984	근로	0.003884	천안	0.003603
근무	0.006392	확산	0.004894	도시	0.003861	인력	0.003586

4.3.2 외국인근로자 의미연결망분석 결과

키워드의 연결 관계 및 패턴을 시각화한 의미연결망분석 결과는 Fig. 4와 같다. 또한, 반복적인 상관관계를 파

악하여 적절한의 군집을 형성하기 위해 CONCOR 분석을 한 다음 Fig. 5의 결과를 도출하였다. 각 군집을 주요 인식과 주변인식으로 분류하여 Table 5에 제시하였다.

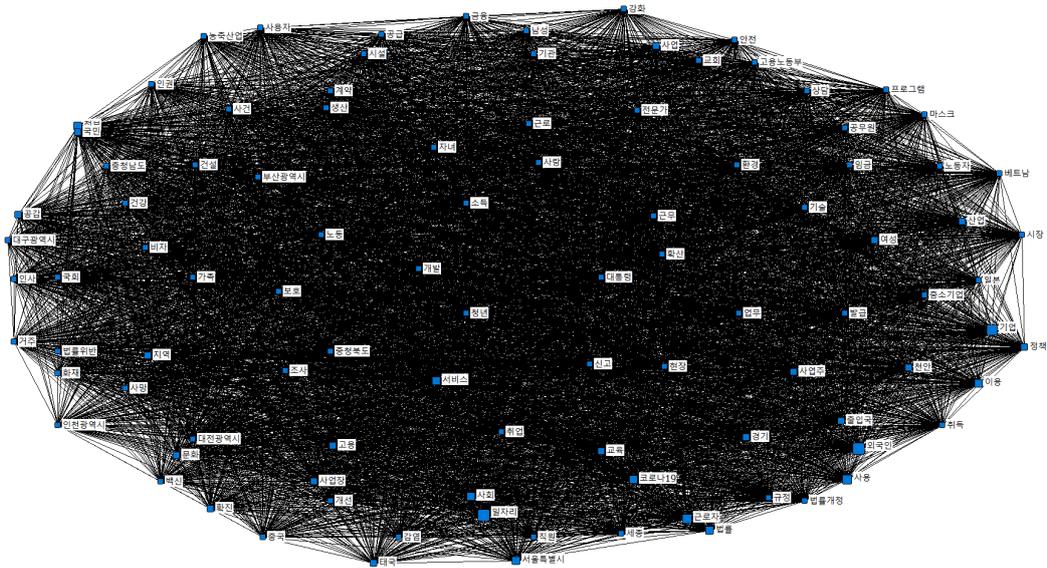


Fig. 4. Semantic Network Analysis of Foreign Workers

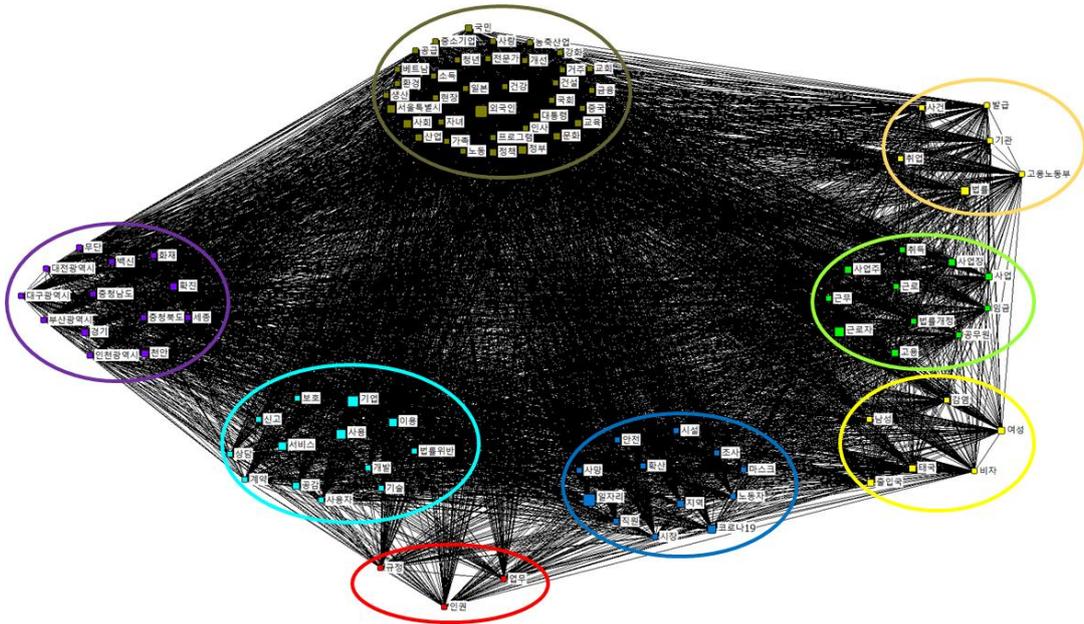


Fig. 5. Visualization of Foreign Workers CONCOR Analysis results

그 결과 첫 번째로 산업, 사회, 생산, 교회, 인사, 농축 산업, 베트남, 건설, 정책, 현장, 서울특별시, 프로그램, 가족, 전문가, 금융, 노동, 외국인, 청년, 정부, 교육, 중국, 사랑, 문화, 일본, 환경, 대통령, 강화, 국회, 중소기업, 건강, 거주, 국민, 공급, 개선, 자녀, 소득, 가족 키워드가 군

집으로 형성되었다. 이를 통해 가장 주요한 이슈로 외국인정책으로 인식하고 있었으며, 이를 주요인식으로 설정하고, '외국인정책 이슈'로 정의하였다.

두 번째로 천안, 대전광역시, 충청남도, 대구광역시, 화재, 세종, 부산광역시, 무단, 확진, 백신, 충청북도, 인

Table 5. Classification of Perceptions of Foreign Workers According to The CONCOR Analysis

Category	Identification name	Word
Major Perception	Foreign policy issues	외국인, 서울특별시, 사회, 정부, 문화, 교육, 국민, 산업, 정책, 중소기업, 중국, 현장, 가족, 건강, 대통령, 환경, 국회, 거주, 일본, 베트남, 프로그램, 개선, 강화, 건설, 금융, 청년, 사망, 소득, 노동, 자녀, 인사, 전문가, 생산, 공급, 교회, 농축산업
Perception of Surrounding	Local community issues	확진, 천안, 경기, 백신, 충청북도, 충청남도, 인천광역시, 대전광역시, 대구광역시, 화재, 세종, 부산광역시, 무단
Perception of Surrounding	Employer's perspective issues	기업, 사용, 서비스, 이용, 공감, 신고, 상담, 보호, 기술, 개발, 법률위반, 계약, 사용자
Perception of Surrounding	COVID-19 issues	일자리, 코로나19, 지역, 시장, 노동자, 직원, 조사, 안전, 확산, 마스크, 사망, 시설
Perception of Surrounding	Working environment issues	근로자, 사업, 고용, 사업주, 사업장, 법률개정, 근무, 취득, 임금, 공무원, 근로
Perception of Surrounding	Legal issues	법률, 취업, 발급, 사건, 기관, 고용노동부
Perception of Surrounding	Immigration issues	출입국, 태국, 여성, 감염, 비자, 남성
Perception of Surrounding	Human rights issues	규정, 업무, 인권

천광역시 키워드가 군집으로 형성되었다. 이를 통해 지역 사회로써 이슈가 존재함을 확인할 수 있었으며, 이를 주변인식으로 설정하고, '지역사회 이슈'로 정의하였다.

세 번째로 계약, 보호, 사용자, 상담, 법률위반, 서비스, 신고, 공감, 기술, 사용, 기업, 이용, 개발 키워드가 군집으로 되었다. 이를 통해 사업주 관점의 이슈가 존재함을 확인할 수 있었으며, 이를 주변인식으로 설정하고, '사업주 관점 이슈'로 정의하였다.

네 번째로 지역, 마케팅, 확산, 사망, 시설, 시장, 안전, 직원, 조사, 일자리, 노동, 코로나19 키워드가 군집으로 형성되었다. 이를 통해 지역사회 이슈가 존재함을 확인할 수 있었으며, 이를 주변인식으로 설정하고, '고용 이슈'로 정의하였다.

다섯 번째로 공무원, 근로, 근무, 근로자, 사업주, 고용, 임금, 법률개정, 사업장, 사업 키워드의 군집으로 형성되었다. 이를 통해 근로환경에 대한 이슈가 존재함을 확인할 수 있었으며, 이를 주변인식으로 설정하고, '근로환경 이슈'로 정의하였다.

여섯 번째로 사건, 기관, 고용노동부, 취업, 법률, 발급 키워드가 군집으로 형성되었다. 이를 통해 법적 이슈가 존재함을 확인할 수 있으며, 이를 주변인식으로 설정하고, '법적 이슈'로 정의했다.

일곱 번째로 출입국, 태국, 비자, 감염, 여성, 남성 키워드는 '출입국 이슈'로 나뉘고, 여덟 번째로 인권, 업무, 규정 키워드가 군집으로 형성되었다. 이를 통해 인권과 관련된 이슈로 존재함을 확인할 수 있으며, 주변인식으로 설정하고, '인권 이슈'로 정의했다.

5. 결론 및 한계점

5.1 결론

본 연구의 목적은 외국인근로자에 대한 인식이 어떻게 형성되어 있는지를 확인하는데 있다. 이를 위해 2020년 1월 1일부터 12월 31일까지 1년 동안의 데이터를 수집하여 텍스토프와 Ucinet6.0을 활용해 단순 빈도분석과 TF-IDF 분석, 연결중심성 분석, 의미연결망 분석, CONCOR 분석을 수행하였다. 연구 결과는 제4차 외국인정책기본계획을 수립 시 기초자료로 활용될 수 있을 것이다. 이 글의 시사점은 크게 세 가지이다.

첫째, 외국인근로자 키워드에서 확진, 천안, 경기, 백신, 충청북도, 충청남도, 인천광역시, 대전광역시, 대구광역시, 화재, 세종, 부산광역시, 무단과 같은 지역사회 이슈를 파악할 수 있었다. 이는 코로나19 상황 속에서 지역사회에 거주하고 있는 외국인근로자의 전염병 문제와 지역 내 외국인근로자에 대한 사회통합적 이슈가 소셜플랫폼을 통해 소통되고 있다는 점을 의미한다. 실제로 효율적이고 효과적인 지방자치단체의 정책 수립을 위해 빅데이터를 활용하여 정책 이슈를 발굴하고 활용한 사례가 있으며[27], 외국인근로자에 대한 정책 수립 시, 소셜데이터를 활용하여 지역사회의 이슈를 분석할 필요가 있다.

둘째, 외국인 근로자와 관련하여 코로나19 이슈가 부각되었다. 지난 한 해 코로나19 상황이 지속되면서 사업장 내 외국인 근로자 감염 확산이 우려가 되었기 때문이다. 이 밖에도 외국인 계절노동자가 입국하지 못해 농어촌의 일손이 부족한 상황을 맞이하는 등 코로나19와 외

국민근로자는 유의한 상관관계를 보였다.

셋째, 사업주 관점의 이슈와 근로자 중심의 근로자환경 이슈가 동시에 나타났다. 근로자, 사업장, 근무, 취득, 임금 등에 대한 근로환경 이슈와 함께 기업, 사용, 이용, 신고, 사용자 등을 아우르는 사업주 관점의 이슈가 모두 중요하게 취급되고 있는 것이다. 그러므로 외국인근로자에 관한 인식을 이해하는 데 있어서 주체인 근로자뿐만 아니라 근로자를 고용하는 사용자들의 관점도 반영되어야 할 것이다.

5.2 한계점

이 논문은 기존 선행연구들과 달리 설문방식의 연구와 법적 쟁점을 분석한 사례 연구에서 벗어나, 소셜플랫폼 상에서 외국인근로자에 대한 국민적 인식을 분석한 실증적인 연구라는 데 의의가 있다.

그럼에도 불구하고 본 연구가 가지고 있는 한계점은 트렌드를 분석하기 위해 설정해놓은 기간이 총 1년에 불과하여 외국인근로자에 대한 인식의 흐름을 파악하는데 충분한 시간을 확보하지 못했다는 것이다. 또한, 코로나 19와 같은 특수한 상황으로 인해 방역 문제가 크게 부각되면서 외국인에 대한 전반적인 인식이 코로나19 이슈에 치우친 것으로 보인다. 따라서 코로나19 상황이 개선되는 시점까지도 연구를 확장하여 정확한 트렌드를 분석할 필요가 있다. 또한, 외국인근로자에 대한 집단 외부적인 소셜네트워크 이외에도 외국인근로자들이 주로 이용하는 소셜네트워크서비스를 분석하여 집단 내부적으로 발생하는 다양한 문제점이나 이슈를 파악하는 방식으로 내외부적인 요소를 모두 고려한 연구가 후속되어야 할 것이다.

REFERENCES

- [1] K. T. Oh. (2007). Korean Workers' Perception to Foreign Workers. *Labor Review*, 25, 59-68.
- [2] J. H. Heo, S. Y. Yu & C. J. Lee. (2020). A Qualitative Study of Intergenerational Differences in Perception on Labor Immigration between Korean 20s and 50s Residing in the Metropolitan Area. *Social Science Research Review*, 36(3), 1-29. DOI : 10.18859/ssrr.2020.8.36.3.1
- [3] Y. J. Im. (2012). Comparative Analysis of Media on Migrant Workers in Korea : With an emphasis analysing Chosun Ilbo, Hankyoreh, Kyoungin and Busan Ilbo. *Journal of Communication Science*, 12(4), 419-456.
- [4] T. H. Kim & S. W. Kim. (2019). Social Tendency and Network Analysis of High School Credit System. *Journal of Educational Innovation Research*, 29(2), 225-242. DOI : 10.21024/pnuedi.29.2.201906.225
- [5] D. H. Seol. (2005). Migrant Workers and Human Rights. *Journal of Democracy and Human Rights*, 5(2), 39-77.
- [6] S. T. Roh & N. C. Kim. (2007). The Human Rights of Aliens in Korea : Focusing on the Impact of International Human Rights Law on Their Situation. *LAW REVIEW*, 48(1), 255-288.
- [7] Ministry of Justice Korea Immigration Service. (2018). *Korea Immigration Service Statistics 2017*. <http://www.immigration.go.kr/bbs/immigration/228/440604/artclView.do>
- [8] Ministry of Justice Korea Immigration Service. (2019). *Korea Immigration Service Statistics 2018*. <http://www.immigration.go.kr/bbs/immigration/228/503695/artclView.do>
- [9] Ministry of Justice Korea Immigration Service. (2020). *Korea Immigration Service Statistics 2019*. <http://www.moj.go.kr/bbs/immigration/228/527712/artclView.do>
- [10] Sulova, S., Todoranova, L., Penchev, B., & Nacheva, R. (2017). Using text mining to classify research papers. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 17(21), 647-654. DOI : 10.5593/sgem2017/21/S07.083
- [11] Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). In *KDD 95*, 112-117.
- [12] Y. M. Jung. (2012). *Research in Information Retrieval*, Seoul : Yonsei University Press
- [13] J. H. Lee & S. H. Lee. (2018). A Case Study on Characteristics of Gender and Major in Career Preparation of University Students from Low-income Families: Application of Text Frequency Analysis and Association Rules. *Journal of Digital Convergence*, 16(12), 61-69. DOI : 10.14400/JDC.2018.16.12.061
- [14] N. G. Kim, D. H. Lee & H. C. Choi. (2017). Investigations on Techniques and Applications of Text Analytics. *The Journal of Korean Institute of Communications and Information Sciences*, 42(2), 471-492. DOI : 10.7840/kics.2017.42.2.471
- [15] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523. DOI : 10.1016/0306-4573(88)90021-0
- [16] H. J. Kim, N. O. Jo & K. S. Shin. (2015). Text Mining-Based Emerging Trend Analysis for the Aviation Industry. *Journal of Intelligence and Information Systems*, 21(1), 65-82. DOI : 10.13088/jiis.2015.21.1.65

- [17] Wang, W., & Rada, R. (1998). Structured hypertext with domain semantic. *ACM Transactions on Information Systems (TOIS)*, 16(4), 372-412.
DOI : 10.1145/291128.291132
- [18] H. W. Park & Loet Leydesdorff. (2004). Understanding the KrKwic: A computer program for the analysis of Korean text. *Journal of the Korean Data Analysis Society*, 6(5), 1377-1387.
- [19] S. S. Lee. (2014). A Content Analysis of Journal Articles Using the Language Network Analysis Methods. *Journal of the Korean Society for information Management*, 31(4),49-68.
DOI : 10.3743/KOSIM.2014.31.4.049
- [20] Y. J. Choi & S. H. Kweon. (2014). A Semantic Network Analysis of the Newspaper Articles on Big data. *Journal of Cybercommunication Academic Society*, 31(1), 241-286.
- [21] Y. H. Kim & Y. J. Kim. (2016). *Social Network Analysis*. Seoul : ParkYoungSa
- [22] K. Y. Kwahk. (2014). *Social Network Analysis*. Seoul : Cheong-Ram
- [23] K. J. Han. (2003). The Meaning and Research Agenda in Network Analysis as Social Science Methodology -based on semantic network analysis-. *Research in Social Studies Education*, 10(2), 219-235.
- [24] J. W. Kang & Y. Namkung. (2020). Understanding Consumers' Perceptions of the Fresh-Food Delivery Platform Service Based on Big Data: Using Text Mining and Semantic Network Analysis. *Korean Journal of Hospitality & Tourism* 30(2), 37-52.
DOI : 10.24992/KJHT.2021.2.30.02.37
- [25] J. S. Lee & C. Y. Kim. (2020). Popularization of Marathon through Social Network Big Data Analysis : Focusing on JTBC Marathon. *Journal of the Korea Entertainment Industry Association* 14(3), 27-30.
DOI : 10.21184/jkeia.2020.4.14.3.27
- [26] J. H. Lee, J. M. Lee, J. H. Kim & H. G. Kim. (2017). A Study on the Perception Change in Marine Sports by Social Media Big Data Analysis. *Korean Journal of Sport Management*, 22(1), 31-46.
DOI : 10.31308/KSSM.22.1.3
- [27] Y. J. Kim & D. Y. Kim. (2018). Methodology of Local Government Policy Issues Through Big Data Analysis. *Journal of the Korea Contents Association*, 18(10), 229-235.
DOI : 10.5392/JKCA.2018.18.10.229

하 재 빈(Jae-Been Ha)

[장학원]



- 2016년 8월 : 국민대학교 정보시스템 전공(경영학사)
- 2018년 3월 ~ 현재 : 아주대학교 비즈니스에널리틱스전공 박사수료
- 관심분야 : 텍스트마이닝, 소셜네트워크 분석, 정보화정책
- E-Mail : jbeyond06@ajou.ac.kr

이 도 은(Do-Eun Lee)

[장학원]



- 2011년 2월 : 영남대학교 법학과(법학사)
- 2013년 2월 : 이화여자대학교 국제대학원 국제학과 국제관계 전공(국제학석사)
- 2018년 5월 : George Washington University Law School, International and Comparative Law(LL.M.)
- 2016년 8월 ~ 현재 : 고려대학교 일반대학원 법학과
- 관심분야 : 국제인권법, 이주민 인권, 고용허가제
- E-Mail : favorite6844@hanmail.net