



Topic Modeling Analysis of Franchise Research Trends Using LDA Algorithm*

LDA 알고리즘을 이용한 프랜차이즈 연구 동향에 대한 토픽모델링 분석

Hoe-Chang YANG 양희창¹

Received: September 23, 2021. Revised: November 14, 2021. Accepted: December 05, 2021.

Abstract

Purpose: This study aimed to derive clues for the franchise industry to overcome difficulties such as various legal regulations and social responsibility demands and to continuously develop by analyzing the research trends related to franchises published in Korea. **Research design, data and methodology:** As a result of searching for 'franchise' in ScienceON, abstracts were collected from papers published in domestic academic journals from 1994 to June 2021. Keywords were extracted from the abstracts of 1,110 valid papers, and after pre-processing, keyword analysis, TF-IDF analysis, and topic modeling using LDA algorithm, along with trend analysis of the top 20 words in TF-IDF by year group was carried out using the R-package. **Results:** As a result of keyword analysis, it was found that businesses and brands were the subjects of research related to franchises, and interest in service and satisfaction was considerable, and food and coffee were prominently studied as industries. As a result of TF-IDF calculation, it was found that brand, satisfaction, franchisor, and coffee were ranked at the top. As a result of LDA-based topic modeling, a total of 12 topics including "growth strategy" were derived and visualized with LDAvis. On the other hand, the areas of Topic 1 (growth strategy) and Topic 9 (organizational culture), Topic 4 (consumption experience) and Topic 6 (contribution and loyalty), Topic 7 (brand image) and Topic 10 (commercial area) overlap significantly. Finally, the trend analysis results for the top 20 keywords with high TF-IDF showed that 10 keywords such as quality, brand, food, and trust would be more utilized overall. **Conclusions:** Through the results of this study, the direction of interest in the franchise industry was confirmed, and it was found that it was necessary to find a clue for continuous growth through research in more diverse fields. And it was also considered an important finding to suggest a technique that can supplement the problems of topic trend analysis. Therefore, the results of this study show that researchers will gain significant insights from the perspectives related to the selection of research topics, and practitioners from the perspectives related to future franchise changes.

Keywords: Empathy Leadership, Job Engagement, Team Cohesiveness, Hotel Chef's

키워드: 프랜차이즈, 프랜차이즈 연구동향, TF-IDF, LDA, Topic Modeling

JEL Classification Code: L1, L6, L8, O1, O3

1. 서론

2021년 국가통계포털 (KOSIS)의 프랜차이즈통계에 따르면, 2019년 국내 프랜차이즈 사업체 수는 21만 5천개를 넘었으며, 종사자수 85만 명, 매출액은 74.6조 원을 넘는 것으로 보고되고 있다 (KOSIS, 2021). 프랜차이즈산업의 매출액은 2019년 기준 한국

* This paper was supported by Jangan University Research Grant in 2021

¹ First Author's Affiliation: Assistant Professor, Department of Distribution Management, Jangan University, Hwaseong, Korea, Email: pricezzang@jangan.ac.kr

© Copyright: Korean Distribution Science Association (KODISA)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

GDP 1,919 조 원과 비교하면 3.89%를, 2019 년 고용보험에 가입된 고용 인력 1,386 만 명과 비교하면 약 6.1%를 차지하는 등 국내 산업을 이끄는 중요한 영역을 차지하고 있다. 이처럼 프랜차이즈사업은 2000 년부터 급속한 발전을 거듭한 산업군 중 하나로, 21 세기 최고의 혁신 경영기법으로 평가될 뿐 아니라 막대한 자본이 없어도 빠른 시간 내에 성공할 수 있는 지식산업으로 분류되고 있다 (Kim & Cho, 2020). 특히 어려운 국내환경에서 프랜차이즈사업은 창직과 창업이라는 화두에 적극 대응함으로써 많은 노동인구를 유입할 수 있는 계기를 제공한 중요한 역할을 하였다. 그러나 그런 긍정적인 결과에도 불구하고 최근 프랜차이즈산업은 양적성장으로 인한 다양한 성장통을 앓게 되었으며 (Han, Lee, & Lee, 2018; Park, Lee, & Choi, 2015), 오랫동안 지속된 불경기와 함께 COVID-19 의 장기 창궐에 따른 어려움 (Kwon, Han, & Lee, 2020)뿐 아니라, 가맹본부에 대한 법적 규제 및 사회적 책임 요구 (Hur, Nor, & Lee, 2019) 등으로 인해 산업의 지속가능성에 대한 우려가 대두되고 있는 형편이다.

따라서 이런 어려운 상황을 타개하기 위해서는 다양한 면에서 문제를 해결할 수 있는 방안을 강구해야 한다. 이 중에서도 전문가들의 시각에서 프랜차이즈 산업에 대한 검토 결과를 확인하는 것은 산업전반의 방향을 확인할 수 있을 것으로 기대할 수 있다.

본 연구는 국내 많은 학회들이 출간하고 있는 학술지 중 프랜차이즈에 대한 학술논문에 주목하고 있다. 왜냐하면 학술지는 연구자들이 관심을 두는 다양한 학문 분야의 이론적 관심과 실제적 관심을 비추어주는 거울이며, 이론적 체계와 연구 방법의 방향을 알려주는 등대일 뿐 아니라 다양한 이론적, 사회적 현상들의 변화에 대한 역사적 기록이라 할 수 있기 때문이다 (Kim, Jeong, Kwan, Lee, & Kim, 2011; Yang, 2019, Yang & Yang, 2021). 이런 이유로 많은 연구자는 다양한 분야에서 연구한 자신의 연구 결과물을 논문게재라는 방법을 통해 세상에 알리고자 노력하고 있으며, 학술지는 이런 논문들을 발표하는 장으로의 역할을 하고 있다 (Yang & Yang, 2021). 또한, 각 학회는 보유한 학술지의 연구 동향과 관련된 연구 발표를 독려하고 있다. 이는 학문의 발전이 외부 환경의 변화에 대응하고 적응하여야 하기 때문이기도 하며, 학문의 발전을 위해 이루어져야 할 가장 기초적인 작업이 기존에 이루어진 연구의 성과물을 체계적으로 정리하는 일이기 때문이다 (Moon & Choi, 2018). 이런 과정을 통해 학술지들의 정체성을 확인할 수 있으며, 연구현황과 동향을 객관적으로 파악함으로써 해당 분야의 발전과 최근 관심사, 그리고 미래 방향에 다양한 단서들을 확인할 수 있다 (Moon & Choi, 2018). 그러나 자체 학술지라는 한정된 영역에 대한 체계적 정리는 그 당위성과 의미에도 불구하고 전체적인 관점에서 진단은 불가능한 문제를

내포하고 있다. 왜냐하면 유사한 학술지들끼리도 해당 분야에 대한 관심사가 다르기 때문이다. 따라서 좀 더 포괄적인 연구를 중심으로 연구 동향을 살펴보는 노력은 관련 학술지들이 자신들의 발전 방향을 진단할 수 있는 지표가 될 수 있다. 따라서 본 연구의 주요 목적은 국내 프랜차이즈 산업에 대한 각 학술지의 연구 동향을 확인함으로써 프랜차이즈 산업의 지속가능성을 위한 다양한 단서들을 도출하고자 하는 것이다.

프랜차이즈산업은 유통, 물류와 직접적인 관계가 있는 산업 분야로 실업문제가 심각한 시대적 환경에 적합한 업종으로 노동인구의 흡수 및 경제 활성화를 위해 가장 적합한 산업 분야라 할 수 있다. 따라서 프랜차이즈산업의 과거와 현재에 대한 단서들을 기반으로 도출한 결과물의 점검을 통해 미래의 프랜차이즈산업에 대한 지속가능성을 예측할 수 있다면 현재 가맹본부 또는 가맹점을 운영하고 있거나 지식산업으로 프랜차이즈산업에 도전하는 잠재사업가들의 성장에 도움이 되는 다양한 방안들이 제시될 수 있을 것으로 기대된다. 이를 위해 본 연구에서는 ScienceON (www.scienceon.kisti.re.kr)에서 프랜차이즈로 분석한 논문 초록을 수집하여 키워드를 추출하고, 이를 LDA (Latent Dirichlet Allocation) 분석을 통해 토픽을 선별하여 프랜차이즈산업과 관련된 주요 이슈를 도출하고자 하였다.

본 연구의 결과는 가맹본부, 가맹점뿐 아니라 신진연구자들에게 그동안 전문가들 관심을 가진 연구영역을 확인함으로써 본부 및 가맹점 운영과 프랜차이즈와 관련된 연구영역의 발굴 등과 관련된 다양한 단서들을 도출할 수 있을 것으로 기대된다. 또한 본 연구는 Topic Modeling 의 LDA 분석을 통해 추출된 토픽의 하위요인들을 활용하여 다양한 인과관계모형의 구축을 위한 아이디어를 제시해줄 수 있을 것이며, 인과관계모형을 통한 실증분석이 이루어지는 경우 프랜차이즈산업의 “선택과 집중” 전략을 위한 또 다른 단서들을 제공해줄 것으로 기대한다.

2. 이론적 고찰

2.1. 프랜차이즈

프랜차이즈의 어원은 “자유를 준다”는 의미를 담고 있는 것이라 고 알려져 있다. 프랜차이즈 (또는 프랜차이즈 사업)란 독립적인 가맹본부 (Franchisor)와 가맹점사업자 (Franchisee)가 상호 계약을 통해 이루어지는 사업의 형태로 법에서 정한 프랜차이즈 (가맹사업)에 대한 정의를 살펴보면 다음과 같다. 가맹사업법에 따르면 가맹사업이란 “가맹본부가 가맹점사업자로 하여금 자기의 상표·서비

스표·상호·간판 그 밖의 영업표지 (이하 “영업표지” 라 한다)를 사용하여 일정한 품질기준이나 영업방식에 따라 상품 (원재료 및 부재료를 포함한다. 이하 같다) 또는 용역을 판매하도록 함과 아울러 이에 따른 경영 및 영업활동 등에 대한 지원·교육과 통제를 하며, 가맹점사업자는 영업 표지의 사용과 경영 및 영업활동 등에 대한 지원·교육의 대가로 가맹본부에 가맹금을 지급하는 계속적인 거래관계”로 정의하고 있다. 여기서 가맹부란 가맹사업과 관련해 가맹점 운영권을 부여하는 자이며 가맹점사업자란 이를 부여받은 사업자를 뜻한다. 한편, 한국프랜차이즈협회 (KFA)나 국제프랜차이즈협회 (IFA)의 정의도 가맹본부가 가맹사업자에게 권리나 노하우의 제공을 약정하고 가맹점사업자는 이를 제공받는 대가를 지불하는 것을 약정하는 계약관계로 정의하고 있다. 따라서 프랜차이즈사업은 특정거래에 집중됨으로써 가치가 있는 거래특유투자 (transaction-specific investment)의 한 형태로 가맹점사업자는 가맹본부에 많은 지원을 받을 뿐 아니라 자율성에 제한을 받는 특징을 가지고 있으며, 현지 시장의 특성에 따라 가맹본부의 통제에서 어느 정도 자율성을 제공받기도 한다 (Pizanti & Lemer, 2003).

프랜차이즈 (또는 프랜차이즈산업)에 대한 국내외 연구 (예: Chun, Lee, & Park, 2020; Han & Han, 2020; Hur & Lee, 2021; Kim, Cho, & Kim, 2019; Kim, Youn, & Moon, 2021; Lanfranchi, Strehlau, Borini, & Melo, 2021; Seo & Park, 2014; Shin & Kim, 2021; Wong, 2021)는 업종별로 프랜차이즈기업의 SNS활동, 기업가정신, 가맹본부의 갑질, 구성원의 직업정체성, 가맹점의 심리적 요인 등 다양한 주제들이 발표되고 있다. 또한 연구동향에 대한 연구들도 학술지 분야 (Moon & Choi, 2018; Yang & Yang, 2021), 건설분야 (Nam & Jang, 2021), 인문학 분야 (Chung, 2021) 등 다양한 분야에서 발표되고 있다. 그러나 해양 엔지니어링과 관련된 연구동향을 보고한 Jilani and Banerjee (2021)의 연구 등 산업이나 기술과 관련된 해외 연구들이 일부 보고되고 있으나 프랜차이즈와 관련한 국내외 연구동향 분석은 확인할 수 없었다.

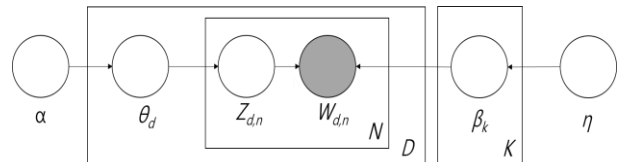
2.2. 토픽모델링

토픽모델링은 텍스트마이닝 기법 중 하나로 비구조화된 텍스트 자료들의 집합에서 의미있는 주제 (토픽)를 추출해주는 확률적 모델 알고리즘이다 (Blei, Ng, & Jordan, 2003; Park & Oh, 2017). 토픽모델링은 방대한 양의 문서집합에서 주요 토픽을 추출하고 각 토픽에 대응되는 문서를 식별하여 제공하며, 하나의 문서가 하나의 토픽으로만 할당되는 일반적인 군집화 (clustering) 기법과는 다르게 하나의 문서가 여러 토픽에 동시에 대응될 수 있기 때문에 현실

세계의 모델링에 보다 적합한 기법으로 평가받고 있다 (Kim, Lee, Choi, & Wong, 2017; Park & Oh, 2017).

토픽모델링의 초기에는 확률적 잠재의미분석 (pLSA; probabilistic Latent Semantic Analysis) 기법이 사용되었으나 본 연구에서 채택한 LDA (Latent Dirichlet Allocation) 분석은 잠재의미분석의 확률분포에 대한 한계점을 보완한 것으로 토픽모델링의 가장 대표적인 방법론으로 활용되고 있다 (Blei et al, 2003).

LDA는 추출된 토픽의 단어 비중과 문서의 토픽 비중이라는 두 가지 변수의 결합확률분포에 따라 문서의 토픽을 찾는 과정으로 두 변수 모두 양의 실수를 요소로 가지며, 모든 요소를 더한 값이 1 이 되는 Dirichlet 분포를 따르는 분석 방법이다 (Blei et al., 2003; Moon, Kim, & Kim, 2019; Yoon & Kim, 2020). LDA는 문서, 단어 등 관찰된 변수를 통해 문맥이나 문서의 구조 등 보이지 않는 변수를 추론하는 방법 (Park & Oh, 2017)으로 토픽에 따라 문서 집합을 재구성할 수 있으며, 데이터, 이미지 및 소셜 네트워크에서 특정한 패턴을 찾는 데 사용될 뿐 아니라 (Blei, 2012), 인터넷 리뷰, 신문기사, 논문 등의 동향을 파악하는 등 잠재 주제를 파악하는데 유용하게 활용되고 있다 (Jo, Kim, & Lee, 2019).



Note) This figure is a quotation from <Figure 1> of Blei (2012).

Figure 1: Graphical model representation of LDA

Figure 1: Graphical model representation of LDA

Figure 1 은 Blei (2012)의 LDA 그래프모델을 인용한 것으로 구체적인 내용은 다음과 같다. 먼저, K 는 결정된 토픽의 개수를 의미하며, α 는 θ 값을 결정하는 파라미터이고, η 는 β 값을 결정하는 파라미터이다 (Park & Oh, 2017). θ 는 문서들의 토픽 비율 (topic proportions)이며, β 는 토픽별 단어인 W 의 생성비율 (per-corpus topic distributions)이고, $Z_{d,n}$ 은 문서 d 의 n 번째 단어의 토픽을 뜻하고, $W_{d,n}$ 은 문서에서 관측되는 변수 (observed variable)로 문서 d 의 n 번째 단어를 뜻한다. 따라서 θ 는 각 문서집합에 대한 주제의 비율 값으로 Dirichlet 분포를 따르며 θ 값에 따라 문서집합 내에 존재하는 단어들의 주제인 Z 가 결정되고, 문서들의 단어의 주제인 Z 와 토픽별 단어의 생성비율인 β 값에 따라 단어 W 가 결정되는 구조이다 (Park & Oh, 2017).

LDA 토픽모델링을 활용한 연구동향 연구로는 교육분야 (Kim & Jhun, 2021; Son & Hwang, 2021, Son, Ahn, Yoo & Lee, 2021), 사회분야 (Lee, 2021) 해외 분야 연구 (Park, 2021) 등 다양한 분야의 연구들이 보고되고 있다.

3. 연구설계

본 연구의 수행을 위해 토픽모델링 중 LDA 기법이 활용되었으며, 토픽과 단어의 관계를 전반적으로 살펴볼 수 있는 데이터 시각화를 위해 LDAvis (Sievrt & Shirley, 2014)가 활용되었다. LDAvis 는 각 토픽과 토픽 내 단어를 중요도에 따라 순위화하여 해당 문서집합에서 주요한 토픽과 단어를 쉽게 파악할 수 있는 웹 기반 토픽모델링 시각화 도구이다. 분석은 R Package 4.1.1 이 이용되었으며, 연구 절차는 Figure 2 와 같다.

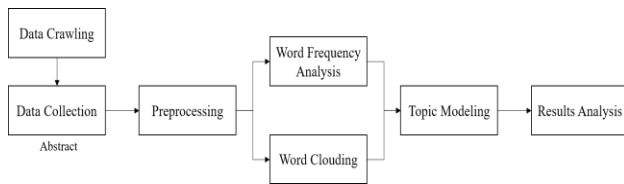


Figure 2: Research Process

3.1. 데이터 수집

본 연구의 수행을 위해 2021 년 6 월 15 일 기준 ScienceON (구 NDSL: National Digital Library)에서 국내학술논문을 대상으로 “프랜차이즈”, “프랜차이즈산업”, “가맹사업”, “가맹본부”, “가맹점”의 5 개 단어를 검색한 결과 <표 1>과 같이 검색되었다. 본 연구는 학술지 초록을 대상으로 분석하고자 하였다. 이는 학술지 초록은 주제어를 모두 담고 있을 뿐 아니라 연구와 관련된 중요한 내용을 모두 함축한 것이기 때문이다. 또한 국문초록과 영문초록의 탑재 빈도를 확인한 결과, 국문초록의 빈도가 영문초록에 비해 40%가 되지 않는 것으로 나타나 검색된 1,142 개의 영문초록이 있는 논문이 확보되었으며, 이 중 ScienceON 의 관리번호가 다른 동일 논문이 32 개 있는 것으로 확인되어 분석 대상 논문은 총 1,110 편이다.

3.2. 전처리 (Preprocessing)

전처리 (preprocessing)는 데이터 분석 전에 데이터를 분석하기 좋은 형태로 만드는 과정으로 (Kim, Na, & Park, 2021), 특수기호, 숫자, 불용어 등을 제거함으로써 양질의 데이터를 확보하는 과정이다.

본 연구에서는 ScienceOn 에서 검색한 프랜차이즈 (franchise)와 초록 (abstract)에 빈번히 등장하는 Statistical, Regression, Correlation 과 같은 통계관련 단어들과 TF-IDF (Term Frequency-Inverse Document Frequency) 분석을 통해 최빈도 단어 중 TF-IDF 가 낮은 단어들 (예: swim, won, october, return)을 불용어로 설정하고 이를 불용어사전 (stopword dictionary)으로 만들어 삭제될 수 있도록 코딩하였다. 전처리한 결과, 전처리 전 총 문자수는 1,670,179 개, 총 단어 수는 245,153 개였으며, 전처리 후 총 문자수는 595,922 개, 총 단어 수는 90,656 개가 도출되었다.

4. 실증분석

4.1. 키워드 빈도분석 및 Word Clouding

ScienceON 에서 “프랜차이즈”라는 키워드로 검색한 1994 년부터 2021 년 상반기까지의 관련 연구 1,110 편의 논문에 대해 키워드 분석을 한 결과는 아래 Table 2 와 같다. 분석 결과, 프랜차이즈와 관련된 연구들의 경우 사업 (business), 상표 (brand), 가맹본부 (franchisor), 산업 (industry), 고객 (customer), 점포 (store, shop), 시장 (market), 시스템 (system) 등에 대한 대상과 서비스 (service), 만족 (satisfaction), 품질 (quality), 신뢰 (trust), 성과 (performance) 등에 대한 관심과 함께 업종으로는 음식 (food), 커피 (coffee) 등이 눈에 띄게 연구되고 있는 것으로 나타났다.

Table 2: Top 20 keywords

Rank	Keyword	Freq.
1	ship	1425
2	service	1306
3	business	1282
4	brand	1241
5	franchisor	1090
6	satisfaction	1031
7	market	997
8	performance	786
9	system	781
10	intent	695
11	quality	694
12	customer	686
13	industry	681

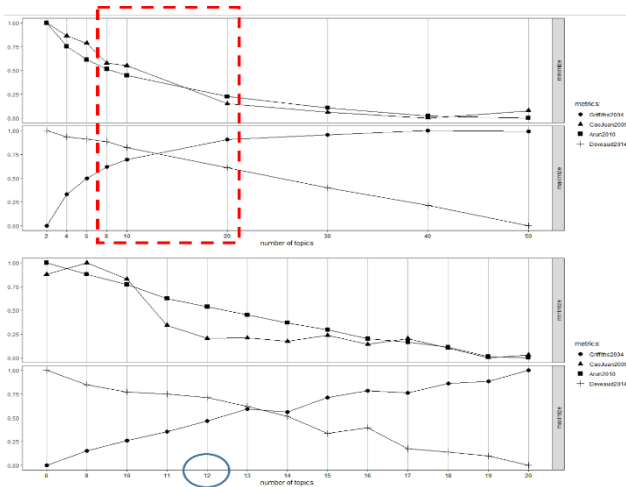


Figure 4: Analysis results to determine the number of topics

프랜차이즈와 관련된 연구에서 중심이 되는 토픽과 관련 핵심어 추출을 위해 LDA 기반 토픽 모델링 알고리즘을 활용하여 토픽모델링을 실시하고 각 토픽의 상위 20개 단어를 도출한 후 각 주제별 토픽 명명을 위해 마케팅을 전공한 박사급 전문가 2 인과 함께 검토한 결과는 Appendix 2 와 같다.

<토픽 1>은 비즈니스 (business), 시장 (market), 산업 (industry), 시스템 (system) 기업 (company), 개발 (develop), 성장 (growth), 전략 (strategy) 등이 연관 단어로 구성되어 이를 프랜차이즈기업의 "성장전략 (growth strategy)"이라고 명명하였다. 토픽 1은 전체 IDM (Intropic Distance Map)의 24%를 차지하고 있는데 이런 결과는 비록 프랜차이즈산업이 어려움을 겪고 있지만, 여전히 이를 극복하기 위한 성장 가능성에 대한 연구자들의 기대와 관심이 반영된 것이라 할 수 있다. <토픽 2>는 "가맹본부 공정성 (Franchisor's Fairness)"으로 가맹본부 (franchisor), 신뢰 (trust), 지원 (support), 공정성 (justice, fairness) 등이 연관 단어로 구성되어 있다. <토픽 3>은 "성과와 갈등구조 (Performance and Conflict Structure)"로 명명되었으며, 갈등 (conflict), 통제 (control), 성과 (performance), 권력 (power), 역할 (role) 등의 연관 단어로 구성되어 있다. <토픽 4>는 "소비경험 (consumption experience)"으로 충성도 (loyalty), 태도 (attitude), 소비 (consume, custom), 구매 (purchase), 경험 (experience) 등의 연관 단어로 구성되어 있다. <토픽 5>는 "규제 (regulation)"로 명명되었으며, 계약 (contract), 법 (law), 합법 (legal), 공정 (fair), 수용 (agree), 보호 (protect), 권리 (right) 등의 연관 단어로 구성되어 있다. <토픽 6>은 "공헌 및 충성도 (contribution & loyalty)"로 명명되었으며, 고객 (customer), 품질 (quality), 만족

(satisfaction), 충성도 (loyalty), 재방문 (revisit) 등의 연관 단어로 구성되어 있다. <토픽 7>은 "브랜드 이미지 (brand image)"로 명명되었으며, 브랜드 (brand), 디자인 (design), 형태 (type), 색깔 (color, colors), 스토리텔링 (storytell) 등의 연관 단어로 구성되어 있다. <토픽 8>은 "지불방식 (payment method)"으로 카드 (card, cards), 신용 (credit), 결제 (pay), 모바일 (mobile), 현금 (cash) 등의 연관 단어로 구성되어 있다. <토픽 9>는 "조직문화 (organizational culture)"로 조직 (organization), 직업 (job), 기업가정신 (entrepreneurship, entrepreneurial), 문화 (culture) 등의 연관 단어로 구성되어 있다. <토픽 10>은 점포 (store), 영역 (area), 입지 (loc), 지역 (local), 시설 (facilities) 등의 연관단어로 구성되어 있어 "상권 (commercial area)"으로 명명되었다. <토픽 11>은 정보 (inform), 온라인 관계망 (SNS, social), 미디어 (media), 네트워크 (network), 사이트 (site), 광고 (advertis), 구전 (wom), 대중 (mess) 등의 연관 단어로 구성되어 프랜차이즈 기업의 "e-비즈니스 (e-business)"로 명명하였다. 마지막으로, <토픽 12>는 "뷰티산업 (beauty industry)"으로 아름다움 (beauty), 가게 (salon), 기술적인 (technical), 헤어 (hair) 등의 연관 단어로 구성되어 있다.

위의 토픽모델링 결과를 LDAvis로 시각화한 결과는 <Figure 5>에 제시되었다. Sievert and Shirley (2014)가 개발한 LDAvis 는 Distinctiveness 와 Saliency 라는 측정 방법을 사용하여 개발한 Trumite (Chaug, Manning, & Herr, 2012)의 한계점을 극복하기 위한 것으로, 토픽과 단어의 관계를 전반적으로 살펴볼 수 있으며, 각 토픽과 토픽 내 단어를 중요도에 따라 순위화하여 해당 문서집합에서 주요한 토픽 및 단어를 쉽게 파악할 수 있는 웹 기반 토픽모델링 시각화도구이다 (Park & Oh, 2017). LDAvis 는 왼쪽에 각 토픽의 연관성과 해당 토픽 내 용어가 전체 토픽에서 전반적으로 사용되는 정도 (prevalence)를 파악할 수 있는 IDM (Intropic Distance Map)이 있으며, 오른쪽에는 전체 토픽에서 가장 중요한 용어들을 보여주는 "Top-30 Most Salient Terms"와 IDM 에서 특정 토픽을 클릭하면 해당 토픽에서 가장 주요한 용어들을 보여주는 "Top-30 Most Relevant Terms Topic N"기능이 있다. 이를 통해 토픽 내의 가장 중요한 단어들이 무엇인지를 쉽게 확인할 수 있으며, 토픽 내의 단어들을 살펴봄으로써 각 토픽이 가진 의미를 파악할 수 있을 뿐 아니라 토픽과 단어 사이의 관계를 통해 새로운 통찰을 얻을 수 있는 장점이 있다 (Park & Oh, 2017).

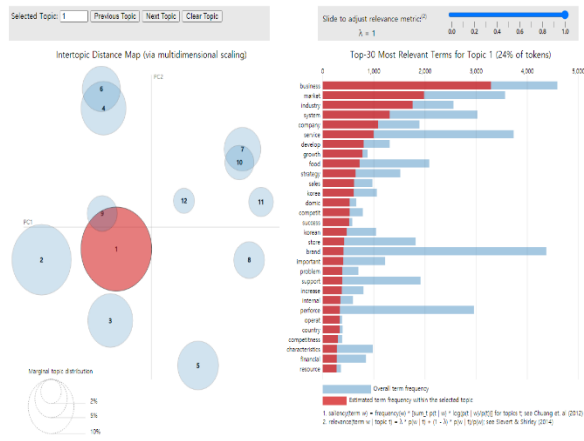


Figure 5: Results of LDA topic modeling visualization

Figure 5에서 보는 바와 같이 토픽 1 (성장전략: growth strategy)과 토픽 9 (조직문화: organizational culture), 토픽 4 (소비경험: consumption experience)와 토픽 6 (공헌 및 충성도: Contribution and Loyalty), 토픽 7 (브랜드 이미지: brand image)과 토픽 10 (상권: commercial area)의 영역이 상당히 중첩되어 있는 것을 알 수 있다. 이런 결과는 프랜차이즈산업의 성장전략은 조직 내부적인 차원과 외부적 상황이 동시에 고려되고 있으며, 소비자의 긍정적인 경험이 공헌도 및 충성도와 연결되고 있으며, 브랜드 이미지와 상권과의 연계성이 반영되고 있음을 의미하는 것이다.

4.3. TF-IDF 를 활용한 키워드 트렌드 분석

토픽 트렌드 분석을 위한 접근 방법으로는 Part and Kim (2019)의 연구에서 활용한 토픽 시계열분석, Cho and Woo (2019)의 연구에서 활용한 특정 연도들에 대한 트렌드 분석 등이 확인되었다. 그러나 이런 분석의 유효성에 대한 점검이 필요하다. 왜냐하면 비지도학습 기반의 LDA 알고리즘을 이용하여 도출된 집단별로 추출된 토픽의 키워드들은 토픽으로 구분 시 중복 할당되기 때문에 각 연도군이나 집단군별로 도출된 토픽별 키워드들은 집단 간에 동일하지 않을 수 있다는 문제가 있기 때문이다. 따라서 본 연구에서는 분석 대상인 전체 논문들의 초록을 통해 도출된 TF-IDF 상위 20 개의 키워드를 대상으로 선정하고, 연도군을 구분한 후, 선정된 각 키워드의 해당 연도군별 TF-IDF 값을 다시 구한 후, 이를 연도군별로 나열하고 추세선 분석을 통해 상승 가능성이 있는 키워드와 하락 가능성이 있는 키워드로 구분하였다. 이는 상승 가능성이 있는 키워드는 향후에도 연구에 중요하게 이용될 주요

키워드로 판단할 수 있으며, 하락 가능성이 있는 키워드는 향후 연구에서 상대적으로 덜 중요하게 이용될 것으로 예상할 수 있기 때문이다. 이를 위해 1994년~2009년 (논문 218편), 2010년~2012년 (논문 225 편), 2013년~2015년 (논문 244 편), 2016년~2018년 (논문 221 편) 및 2020년~2021년 6월 (논문 202 편)의 5개 연도군으로 구분하고 위의 <Table 3>에서 제시한 TF-IDF 기준 상위 10개 키워드와 고객 (customer) 등 11위부터 20위까지의 키워드에 대한 연도군별 TF-IDF 분석을 통해 연도군별 결과가 <Figure 6>에 제시되었다.

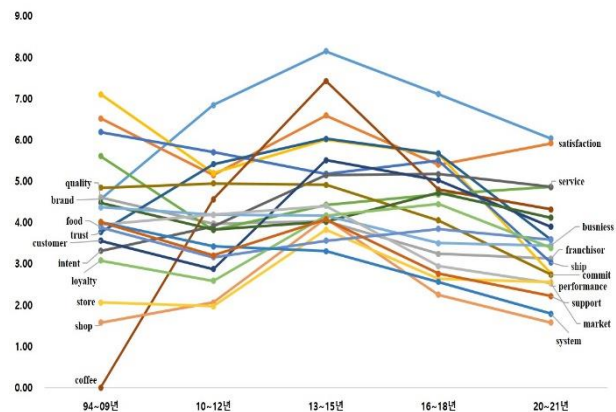


Figure 6: Trend analysis results for TF-IDF top 20 keywords

Figure 6의 좌측에 게재된 품질 (quality), 브랜드 (brand), 음식 (food), 신뢰 (trust) 등 10개 키워드는 전반적으로 상승하고 있는 추세를 보여 프랜차이즈 연구에서 중요하게 다루어지는 키워드라 할 수 있으며, 우측에 게재된 만족 (satisfaction), 서비스 (service) 가맹본부 (franchisor) 등의 10개 키워드는 좌측에 제시한 키워드에 비해 하락하고 있는 추세를 보여 상대적으로 그 중요성이 덜한 것으로 판단할 수 있다.

5. 결론

본 연구는 실업문제가 심각한 우리나라 상황에서 노동인구를 흡수하고 경제 활성화를 위해 가장 적합한 산업분야인 프랜차이즈산업이 다양한 법적 규제 및 사회적 책임요구 등의 어려움을 극복하고 지속적으로 발전할 수 있는 단서를 도출하고자 하였다. 이를 위해 본 연구는 1994년부터 2021년 6월까지 발표된 학술지를 대상으로 프랜차이즈로 검색한 1,110개 유효논문의 초록을 수집하여 키워드를 추출한 후 R 패키지를 활용하여 키워드 분석, TF-IDF 분석 및 LDA 알고리즘을 활용한 토픽모델링과 함께

TF-IDF 상위 20 개 단어들의 연도군별 트렌드 분석을 실시하였다. 분석 결과는 다음과 같다.

첫째, 키워드 빈도분석 결과, 프랜차이즈와 관련된 연구들의 경우 사업 (business), 상표 (brand), 가맹본부 (franchisor), 산업 (industry), 고객 (customer), 점포 (store, shop), 시장 (market), 시스템 (system) 등에 대한 대상과 서비스 (service), 만족 (satisfaction), 품질 (quality), 신뢰 (trust), 성과 (performance) 등에 대한 관심과 함께 업종으로는 음식 (food), 커피 (coffee) 등이 눈에 띄게 연구되고 있는 것으로 나타났다.

둘째, TF-IDF 연산 결과 <표 3>에 제시한 것과 같이 빈도 수가 높은 상위 TF 단어들에 비해 브랜드 (brand), 만족 (satisfaction), 가맹본부 (franchisor), 커피 (coffee) 등이 상위에 위치하는 것으로 나타났다. 이런 결과는 키워드의 빈도에 비해 TF-IDF 가 높은 키워드들이 문서 내에서 상대적으로 더 중요한 역할을 하는 것이기 때문에 관심도가 더 높은 분야라는 것을 의미한다.

셋째, LDA 기반 토픽모델링 결과, 총 12 개의 토픽이 도출되었다 (<Table 4> 참조). 이 중에서 "성장전략 (growth strategy)"이라고 명명한 토픽 1은 비즈니스 (business), 시장 (market), 산업 (industry), 시스템 (system) 기업 (company), 개발 (develop), 성장 (growth), 전략 (strategy) 등의 연관 단어들로 구성되어 있으며, 전체 IDM 중 약 24%를 차지하며 <Figure 5>의 LDAvis 시각화 결과에서 보듯이 토픽 9 (조직문화: organizational culture) 와 상당히 중첩된 것을 알 수 있었다. 또한 12 개의 토픽 중 토픽 4 (소비경험: consumption experience) 와 토픽 6 (공헌 및 충성도: Contribution and Loyalty), 토픽 7 (브랜드 이미지: brand image)과 토픽 10 (상권: commercial area)의 영역이 상당히 중첩되어 있는 것을 알 수 있다. 이런 결과는 프랜차이즈산업의 성장전략은 조직 내부적인 차원과 외부적 상황이 동시에 고려되고 있으며, 소비자의 긍정적인 경험이 공헌도 및 충성도와 연결되고 있음과 함께 브랜드 이미지와 상권과의 연계성이 반영되고 있음을 의미하는 것이다.

넷째, IDM 비율을 볼 때 프랜차이즈 연구의 40.9%가 "성장전략 (growth strategy)", "가맹본부 공정성 (Franchisor's Fairness)" 분야에 집중되어 있으며, "성과와 갈등구조 (Performance and Conflict Structure)", "소비경험 (consumption experience)"까지 더하면 약 60.5%의 비중을 차지하는 것을 확인할 수 있었다. 이런 결과는 현재까지 관심을 가진 분야에 대한 분류로 볼 수 있으나, 한편으로는 좀 더 다양한 분야의 연구를 통해 프랜차이즈산업이 지속적으로 성장할 수 있는 단서들을 찾아야 한다는 것으로도 해석할 수 있다.

마지막으로, TF-IDF 가 높은 상위 20 개의 키워드를 5 개 연도군으로 구분하여 각각의 TF-IDF 를 연산하여 키워드 트렌드 분석을 실시한 결과, 품질 (quality), 브랜드 (brand), 음식 (food),

신뢰 (trust) 등 10 개 키워드가 전반적으로 상승하고 있으며, 만족 (satisfaction), 서비스 (service) 가맹본부 (franchisor) 등의 10 개 키워드는 하락하고 있는 추세를 보여 상대적으로 그 중요성이 덜한 것으로 판단할 수 있다. 이런 결과 또한 <Figure 6>에서 제시된 품질 (quality), 브랜드 (brand), 음식 (food), 신뢰 (trust), 고객 (customer), 관심의도 (intent), 충성도 (loyalty), 점포 (store, shop), 커피 (coffee) 등의 키워드가 앞으로도 연구주제로 더 활용될 것으로 예상된다.

본 연구의 이론적, 실무적 시사점은 먼저, 프랜차이즈 산업의 관심 방향을 LDA 토픽모델링을 통해 제시한 점을 들 수 있다. 구체적으로 1994 년부터 발표된 프랜차이즈와 관련된 논문의 영문초록을 바탕으로 현재까지 연구되고 있는 프랜차이즈산업에 대한 연구주제는 총 12 개로 구분되는 것이 확인되었다. 이 중에서도 성장전략의 경우 조직 내부적인 차원과 외부적 상황이 동시에 고려되고 있음과 소비자의 긍정적인 경험이 공헌도 및 충성도에 연결되고 있고, 브랜드 이미지와 상권과의 연계성이 반영되는 것을 알 수 있었다. 또한, 토픽 트렌드 분석에 대한 한계점을 보완할 수 있는 TF-IDF 트렌드 분석을 실시한 결과 품질, 브랜드, 음식, 신뢰, 고객, 충성도, 점포, 커피 등에 관한 관심이 높아지고 있음을 확인한 것 또한 실무적인 관점에서 관심있게 검토할 필요가 있다. 마지막으로, 12 개의 토픽으로 구성된 하위 키워드들을 통한 인과관계 연구 가능성의 단서들을 제시한 점은 중요한 시사점으로 들 수 있다. 예를 들어, 토픽 중 성장전략과 충성도 간에 가맹본부의 공정성 (또는 규제)의 매개 또는 조절효과를 확인하는 모형이나, 소비자들의 긍정적 소비경험과 충성도 간에 브랜드 이미지의 효과 모형 등의 구축이 가능할 것이다. 또한 토픽 9 의 조직문화에서 기업가정신과 성과 또는 이직의도의 영향관계 등, 토픽 내의 키워드를 활용한 연구모형의 구축 또한 가능할 것으로 판단된다. 이런 연구결과를 통해 연구자들은 연구주제의 선택과 관련된 관점에서, 실무자들은 향후 프랜차이즈의 변화와 관련된 관점에서 통찰력을 얻을 수 있을 것이다.

이런 시사점에도 불구하고 본 연구는 다음과 같은 한계점이 있다. 첫째, scienceON 을 활용하여 획득한 학술지 데이터가 국내연구로 한정된 것을 들 수 있다. 이는 선진 프랜차이즈산업의 동향을 확인하는데 한계가 있음을 의미하기 때문이다. 따라서 향후 연구에서는 좀 더 다양한 해외연구 결과와의 비교연구 등을 통해 연구자들과 실무자들에게 통찰력을 얻을 수 있도록 할 필요가 있다. 둘째, 본 연구에서 전처리과정에서 어간추출 (stemming)을 포기하고 단어중심의 전처리를 한 것을 들 수 있다. 이는 탐재된 영문초록 자체의 문제인지, R 패키지의 문제인지 확인할 수 없었지만 분석과정에서 지속적으로 어간 및 단어 간의 충돌이

발생하였기 때문에 최대한 식별가능한 방법을 적용하였기 때문이다. 예를 들면, 본문의 perforce (performance로 수정함) 등의 용어들은 식별이 가능한 경우 수정하였으나 ship 등의 단어들은 그 의미를 확인할 수 없었다. 따라서 향후 연구에서는 Python 등을 통해 동일한 절차로 분석해 볼 필요가 있다. 마지막으로, 토픽 수 결정과 토픽 명칭부여에 있어 연구자의 자의성을 최소화하고자 하였음에도 불구하고 객관성을 확보하는 데는 한계가 있다. 따라서 향후 연구에서는 본 연구에서 적용한 로그우도 추정치 (Log Likelihood Expectation)를 활용한 토픽 수 결정방법 외에도 응집도 (Coherence)나 혼잡도 (perplexity)를 함께 적용하여 토픽 수를 결정할 필요가 있으며, 토픽 명칭 또한 좀 더 객관적이고 일반적인 단어로 부여하여 연구자나 실무자들이 쉽게 받아들일 수 있도록 하여야 할 것이다.

References

- Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010, June). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 391-402). Springer, Berlin, Heidelberg.
- Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72 (7-9), 1775-1781.
- Cho, K. W., & Woo, Y. W. (2019). Topic modeling on research trends of industry 4.0 using text mining. *Journal of the Korea Institute of Information and Communication Engineering (JKIICE)*, 23 (7), 764-770.
- Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. *Advanced Visual Interfaces*, 12, 21-25.
- Chun, T. Y., Lee, D. K., & Park, N. H. (2020). The effect of marketing activities on the brand recognition, brand familiarity, and purchase intention on the SNS of franchise companies. *Journal of Asian Finance, Economics and Business*, 7 (11), 955-966.
- Chung, E. K. (2021). An investigation on digital humanities research trend by analyzing the papers of digital humanities conferences. *Journal of the Korean Society for Library and Information Science*, 55 (1), 393-413.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, 17, 61-84.
- Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *PNAS*, 1 (101), 5228-5235.
- Han, S. H., Lee, Y. K., & Lee, J. G. (2018). Effects of organizational justice on emotions, job satisfaction, and turnover intention in franchise industry. *Korean Journal of Franchise Management*, 9 (2), 7-16.
- Han, Y. W., & Han, S. H. (2020). Effect of entrepreneurial passion and positive emotions on customer responses: Mixed methods approach. *Journal of Distribution Science*, 18 (5), 35-47.
- Hur, S. B., & Lee, Y.-K. (2021). The effect of franchisors' gapjil on economic satisfaction, social satisfaction, and recontract intention. *Korean Journal of Franchise Management*, 12 (2), 35-49.
- Hur, S. B., Nor, Y. S., & Lee, D. (2019). The impact of franchisor's economic and philanthropic CSR on franchisees' economic satisfaction, social satisfaction, and loyalty. *Korean Journal of Franchise Management*, 10 (3), 25-35.
- Jo, H. I., Kim, J. W., & Lee, B. G. (2019). A study on research trends of blockchain using LDA topic modeling: Focusing on United States, China, and South Korea. *Journal of Digital Contents Society*, 20 (7), 1,453-1,460.
- Jilani, G., & Banergee, S. (2021). Research trend of marine engineering in India and China during 2010-2019. *Qualitative and Quantitative Methods in Libraries*, 10 (1), 129-145.
- Kim, C. B., & Cho, K. R. (2020). A study on the factors affecting the success factors of franchise companies to corporate performance: The moderating effect of entrepreneur capacity. *Korea Journal of Business Administration*, 33 (9), 1671-1694.
- Kim, H. W., & Jhun, Y. S. (2021). Analysis of trends in science gifted education using topic modeling. *Journal of Korean Elementary Science Education*, 40 (3), 283-294.
- Kim, J. Y., Na, H. S., & Park, K. H. (2021). Topic modeling of profit adjustment research trend in Korean accounting. *Journal of Digital Convergence*, 19 (1), 125-139.
- Kim, K. S., Cho, S. H., & Kim, S. H. (2019). A study on effects of franchise distribution employees on occupational identity. *Journal of Distribution Science*, 17 (9), 83-92.
- Kim, K. W., Jeong, J. J., Kwan, H. Y., Lee, Y. J., & Kim, C. K. (2011). Analysis of research in the Korean journal of counseling and psychotherapy (2000~2009). *Korean Journal of Counseling and Psychotherapy*, 23 (3), 521-542.
- Kim, N. G., Lee, D. H., Choi, H. C., & Wong, W. X. S. (2017). Investigations on techniques and applications of text analytics. *Journal of Korean Institute of Communications and Information Sciences*, 42 (2), 471-492.
- Kim, S. Y., Youn, S. J., & Moon, J. S. (2021). The impact of basic psychological needs satisfaction on the performance of the franchisee. *Journal of Distribution Science*, 19 (1), 17-26.
- Kwon, O. S., Han, S. H., Lee, D. H. (2020). SME profitability of trade credit during and after a financial crisis: Evidence from Korea. *Journal of Asian Finance, Economics and Business*, 7 (7), 35-47.
- Lanfranchi, A. G., Strehlau, S., Borini, F. M., & Melo, P. L. d. R. (2021). Does origin matter? The impact of the institutional environment of the origin country on the internationalization of franchise chains. *Multinational Business Review*, 29 (1), 96-115.
- Lee, D. Y., & Yi, H. S. (2021). Exploring methods for determining the appropriate number of topics in LDA: Focusing on perplexity and harmonic mean method. *Journal of Educational Evaluation*, 34 (1), 1-30.
- Lee, S. Y. (2021). An analysis of civil complaints about traffic policing using the LDA model. *Journal of the Korea Institute of Intelligent Transportation Systems*, 20 (4), 57-70.

- Liu, E. (2015). TF-IDF, term frequency-inverse document frequency. [cited 2020 Sept 30]. Available from: https://ethen8181.github.io/machinelearning/clustering_old/tfidf/tfidf.html
- Moon, M. R., Kim, M. C., & Kim, J. W. (2019). A study on the fine dust-related bills in the national assembly: Based on the revision of the special act on fine dust (in Korean). *Hannam Journal of Law & Technology*, 25 (4), 87-115.
- Moon, Y. H., & Choi, J. H. (2018). Content analysis of journal of channel and retailing: Research trends and future directions. *Journal of Channel and Retailing*, 23 (4), 51-73.
- Nam, K. W., Jang, M. H. (2021). Content analysis of vocational research trend in construction field. *Korean Journal of Construction Engineering and Management*, 22 (2), 31-41.
- Nahm, C. H. (2016). An illustrative application of topic modeling method to a farmers diary. *Cross-Cultural Studies*, 22 (1), 89-135.
- Park, J. S., & Kim, C. S. (2019). Research trends analysis of big data: Focused on the topic modeling. *Journal of the Korea Society of Digital Industry and Information Management*, 15 (1), 1-7.
- Park, S. K., Lee, K. W., & Choi, Y. H. (2015). The study of win-win growth and problems between franchisor and franchisee. *Korean Journal of Franchise Management*, 6 (2), 31-49.
- Park, J. H., & Oh, H. J. (2017). Comparison of topic modeling methods for analyzing research trends of archives management in Korea: Focused on LDA and HDP. *Journal of Korean Library and Information Science Society*, 48 (4), 235-258.
- Park, H. Y. (2021). Study on Myanmar research trends based on keyword network analysis (Topic Analysis Using LDA Methodology). *International Commerce and Information Review*, 23 (3), 79-109.
- Pizanti, I., & Lerner, M. (2003). Examining control and autonomy in the franchisor-franchisee relationship. *International Small Business Journal*, 21 (2), 131-159.
- Seo, M. G., & Park, J. H. (2014). Significance and method for a more balanced development of the restaurant franchise business: Focused on performance-sharing examples between franchisor and franchisee. *Journal of Franchise Management*, 5 (1), 65-84.
- Shin, Y. H., & Kim, J. S. (2021). A study on the effect of social franchise characteristics on management performance: Focusing on the mediating effect of franchise autonomy. *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 16 (3), 121-143.
- Sievert, C., & Shirley, K. E. (2014). *LDavis: A method for visualizing and interpreting topics. proceedings of workshop on interactive language learning, visualization, and interfaces*. Baltimore: Maryland.
- Son, T. K., & Hwang, S. H. (2021). Analysis of the research trends of domestic elementary mathematics education using topic modeling. *Journal of Elementary Mathematics Education in Korea*, 25 (1), 61-80.
- Son, T. K., Ahn, D. Y., Yoo, S. Y., & Lee, K. H. (2021). An analysis of research trends of domestic mathematics struggling learners using topic modeling. *Brain, Digital, & Learning*, 11 (1), 1-13.
- Wong, H. S. (2021). Issues and challenges of unregistered franchise business: A Malaysian legal perspective. *Studies of Applied Economics*, 39 (4), 1-10.
- Yang, H. C., & Yang, W. R. (2021). Comparison of research trends in KODISA directly managed journals using keyword analysis. *Journal of Research and Publication Ethics*, 2 (1), 19-24.
- Yang, W. R. (2019). Exploration of research trends in the journal of distribution science using keyword analysis. *International Journal of Industrial Distribution & Business*, 10 (8), 17-24.
- Yoon, S. U., Kim, M. C. (2020). Topic modeling on fine dust issues using LDA Analysis. *Journal of Energy Engineering*, 29 (2), 23-29.
- KOSIS (2021). <https://kosis.kr/search/search.do> ScienceON : <https://scienceon.kisti.re.kr/main/mainForm.do>

Appendix

Appendix1: Search results for domestic academic papers on ScienceON

Division	Franchise	Franchise Industry	Franchise business	Franchisor	Franchisee
Number of Papers	1,324	140	160	207	415

Appendix2: Results of Topic modeling analysis

# of IDM	Ratio	Topic name	Top 20 words in each topic
1	24.0	Growth strategy	business, market, industry, system, company, service, develop, growth, food, strategy, sales, Korea, competit domestic, success, korean, store, brand, important, support
2	16.9	Franchisor's Fairness	ship, franchisor, trust, satisfaction, performance, support, commit, intent, service, orient, quality, affect, market, system reconstruct, justice, industry, food, communication, fairness
3	9.8	Performance and Conflict Structure	ship, franchisor, conflict, system control, performance, loyalty, channel, opportunism behavior, goal power, governance, sturcture, role, distribute, knowledge, level, style, contract
4	9.8	Consumption experience	brand, loaylty, attitude, market, intent, consume, restaurant, customer, trust, service, ship, food, affect, purchase, company, custom, social, experience, csr, corporate
5	8.1	Regulation	franchisor, contract, business, law, trade, act, legal, fair, inform, dispute, agree, transact, protect, korea, problem, system, regulate, right, disclosure, labor
6	7.0	Contribution and Loyalty	service, coffee, customer, quality, satisfaction, shop, custom, intent, store, market, attribute, brand, loyalty, environ, select, price, food, ship, revisit, difference
7	6.5	Brand image	brand, design, bakery, food, ident, consume, type, price, korean, color, colors, visual, local, promot, storytell, interior, product, people, space, form
8	4.6	Payment method	card, credit, pay, service, system, company, fee, mobile, merchants, tax, merchant, real, cash, cards, tie, inform, banks, van, smart, transact
9	4.3	Organizational culture	organizational, job, employee, performance, entrepreneurship, ship, intent, work, commit, behavior, satisfaction, emotal, orient, culture, turnover, entrepreneurial, company psychological, affect, comete,
10	4.1	Commercial area	store, area, retail, commercial, chain, loc, sales, ship, convenience, local, chicken, supermarkets, characteristics, food, coffee, trade, price, popul, content, facilities,
11	2.9	e-business	inform, sns, online, social, media, content, network, search, custom, visit, site, particip, eng, advertis, characteristics, mess, die, wom, consumer, keyword
12	2.1	Beauty industry	beauty, salon, educ, educal, efficie, shop, program, satisfaction, train, business, method, difference, content, system, technical, hair, level, improve, industry, service