

Analysis of eSports Industry Trends Using TF-IDF and LDA Topic Modeling

Tae-Min Lee*, Jeon-Young Kong**, Woong Kim***

*Student, Dept. of Digital Twin Software, Dongyang University, Dongducheon, Korea

**Professor, Dept. of eSports, Dongyang University, Dongducheon, Korea

***Professor, Dept. of Digital Twin Software, Dongyang University, Dongducheon, Korea

[Abstract]

This study analyzed major discourses in the esports industry using TF-IDF and LDA topic modeling. The dataset comprised 40 KIPRIS patents, 289 KeSPA press releases, and 171 Global Esports news articles. Preprocessing included regular-expression-based token extraction, stop-word removal, orthographic normalization, and compound-term preservation. English candidate terms were mapped to representative Korean concepts, and source-specific mean TF-IDF scores were L1-normalized and combined with equal weights to identify key terms. For LDA, the 500 original documents were retained as individual units, with adjustments for document length and source imbalance. Topic models containing two to ten topics were evaluated using coherence, perplexity, diversity, and stability. The analysis identified three source-associated topics: technology and data services, domestic leagues and talent development, and global cooperation and governance. This study contributes by comparing esports-industry discourses across multiple sources while limiting its interpretation to the collected dataset.

▶ **Key words:** eSports, TF-IDF, LDA, Topic Modeling, Web Crawling, Text Mining

[요 약]

본 논문은 TF-IDF와 LDA 토픽 모델링을 활용하여 e스포츠 산업의 주요 담론을 분석하였다. KIPRIS 특허 40건, KeSPA 보도자료 289건, Global Esports 뉴스 171건을 수집하여 정규표현식 기반 토큰 추출, 불용어 제거, 표기 정규화 및 복합어 보존을 수행하였다. 영문 후보 어휘를 한국어 대표 개념으로 매핑한 후, 자료원별 평균 TF-IDF를 L1 정규화하고 동일 가중치로 결합하여 핵심 키워드를 도출하였다. LDA에서는 500개의 원문을 개별 분석 단위로 사용하고 문서 길이와 자료원 불균형을 보정했으며, 토픽 수 2~10개에 대한 일관성, 퍼플렉시티, 다양성 및 안정성을 비교하였다. 그 결과 기술·데이터 서비스, 국내 리그·인재 양성, 글로벌 협력·거버넌스의 세 가지 주요 토픽이 자료원별로 뚜렷하게 나타났다. 본 연구는 서로 다른 자료원에서 관찰된 e스포츠 산업 담론을 다각적으로 분석하되, 해석 범위를 수집자료에 한정했다는 점에서 의의가 있다.

▶ **주제어:** e스포츠, TF-IDF, LDA, 토픽 모델링, 웹 크롤링, 텍스트 마이닝

- First Author: Tae-Min Lee, Corresponding Author: Woong Kim
- *Tae-Min Lee (darky0604@naver.com), Dept. of Digital Twin Software, Dongyang University
- **Jeon-Young Kong (jycong@outlook.kr), Dept. of eSports, Dongyang University
- ***Woong Kim (woong2214@dyu.ac.kr), Dept. of Digital Twin Software, Dongyang University
- Received: 2026. 07. 15, Revised: 2026. 08. 18, Accepted: 2026. 08. 24.

I. Introduction

e스포츠는 전자적 시스템을 매개로 선수와 팀의 경기 수행, 관전자 경험, 중계 및 팬덤 활동이 형성되는 경쟁적 게임 문화로 정의될 수 있다[1]. 초기에는 특정 게임 커뮤니티와 온라인 방송을 중심으로 성장하였으나, 최근에는 프로리그, 국제대회, 스트리밍 플랫폼, 스폰서십, 교육 프로그램, 데이터 분석 서비스가 결합된 복합 산업으로 확장되고 있다[2][3]. 특히 e스포츠는 전통 스포츠와 유사한 경쟁 구조, 팀 단위 운영, 팬 기반 수익 모델을 갖추면서 스포츠 산업과 디지털 콘텐츠 산업 사이의 융합 영역으로 논의되고 있다[4].

선행연구에서는 e스포츠를 스포츠의 정의 안에 포함할 수 있는지, 프로게이머를 선수로 볼 수 있는지, 관람 동기와 팬덤 구조가 전통 스포츠와 어떠한 차이를 갖는지에 대해 지속적으로 논의되었다[5]. 또한 e스포츠는 단순한 개인 게임 플레이가 아니라 팀 조직, 전략 수행, 의사소통, 팬덤 참여, 플랫폼 기반 소비가 동시에 나타나는 사회적 실천으로 이해될 필요가 있다[6]. 이러한 관점에서 e스포츠는 게임 이용 행태의 하위문화가 아니라 미디어, 스포츠, 기술, 교육, 지역 산업이 연결되는 독립적인 연구 대상으로 발전하고 있다[7][8].

산업적 측면에서도 e스포츠는 리그 운영과 방송 중계 중심의 구조를 넘어 글로벌 거버넌스와 국제 종합대회 편입으로 제도화되고 있다. 2022 항저우 아시안게임(2023년 개최)에서 e스포츠가 처음으로 정식 메달 종목으로 채택되어 치러졌으며, 이는 e스포츠가 국제 스포츠 체계 안에서 공식성을 확보하는 중요한 계기가 되었다[9][10]. 또한 Global Esports Federation과 같은 국제 조직은 회원국 네트워크, 글로벌 대회, 교육·개발 프로그램을 확대하면서 세계 단위의 e스포츠 생태계 형성에 기여하고 있다[11]. 국내에서도 KeSPA를 중심으로 선수, 팀, 대회, 지역 리그, 아카데미, 국가대표 체계와 관련한 보도자료가 지속적으로 축적되고 있어 제도화 흐름을 파악할 수 있는 중요한 자료원이 된다[12].

그러나 e스포츠 산업은 게임 종목, 리그 구조, 기술 서비스, 팬덤 문화, 교육 체계가 빠르게 변화하기 때문에 단일 자료원만으로 산업의 방향성을 해석하는 데 한계가 있다. 특히 자료는 기술적 응용과 서비스 개발 방향을 보여주지만, 실제 리그 운영이나 국제 제도화 흐름을 충분히 설명하지 못한다. 반대로 협회 보도자료나 뉴스 자료는 제도, 운영, 행사 중심의 동향을 보여주지만, 기술적 변화나 서비스 구현 수준을 체계적으로 파악하기 어렵다. 따라서

특히, 국내 기관 보도자료, 글로벌 뉴스 자료를 결합한 다중 자료원 기반 분석이 필요하다.

텍스트 마이닝은 대규모 비정형 문서에서 반복적으로 등장하는 핵심 단어와 잠재 주제를 추출하는 데 유용하다[13][14]. 그중 TF-IDF는 특정 문서 집합에서 상대적으로 중요한 단어를 식별할 수 있는 대표적인 가중치 기법이며[15][16], LDA는 문서와 단어 사이의 확률적 관계를 기반으로 잠재 토픽을 도출하는 토픽 모델링 방법이다[17][18]. 최근에는 특히, 뉴스, 논문, 소셜미디어 데이터에 TF-IDF와 LDA를 적용하여 기술 동향, 산업 이슈, 정책 의제를 분석하는 연구가 확대되고 있다[19].

이러한 문제의식에 따라 KIPRIS 특허 자료, KeSPA 보도자료, Global Esports 뉴스 자료를 수집하고 TF-IDF 및 LDA 토픽 모델링을 적용하여 e스포츠 시장의 핵심 동향을 도출하고자 한다. 본 연구의 차별점은 기존의 토픽 모델링 기반 산업 동향 연구가 특허 또는 뉴스 등 단일 자료원에 의존하였던 것과 달리, 본 연구는 기술 자료(특허), 국내 산업 운영 자료(협회 보도자료), 글로벌 뉴스 자료를 통합 분석함으로써 기술, 제도, 글로벌 동향을 입체적으로 파악한다는 점에서 차별성을 가진다. 또한 자료원별 TF-IDF 상위 키워드를 통합하여 중복 키워드를 정리한 뒤 토픽 모델링을 수행함으로써, 자료원 간 공통 담론과 자료원 고유 담론을 동시에 반영하였다. 이후 도출된 토픽을 글로벌 협력·거버넌스, 기술·데이터 서비스, 국내 리그·인재 양성이라는 산업적 관점에서 해석한다.

II. Preliminaries

1. Related works

초기의 e스포츠 연구는 주로 개념 정의와 스포츠성 논의에서 출발하였다. e스포츠는 전자적 시스템을 통해 이루어지는 스포츠 형태로 설명되었으며, 관람 동기와 함께 놀이, 경쟁, 기술, 신체성, 제도화 등의 요소를 기준으로 기존 스포츠 범주에 포함될 수 있는지가 논의되었다[20]. 이에 일부 연구자들은 e스포츠가 단순한 게임 활동을 넘어 스포츠 매니지먼트 분야의 교육 및 연구 대상으로 다루어질 필요가 있다고 보았다.

이후 연구는 사회문화적 특성, 팬덤, 팀 운영, 관람 동기, 산업 구조 등으로 확장되었다. 프로게이밍의 전문화 과정이 분석되면서 e스포츠가 게임 문화의 일부에서 출발해 전문 직업과 산업 구조를 갖춘 분야로 발전해 왔음이 설명되었다. 또한 e스포츠는 개인의 게임 플레이가 아니라

사회적 실천이자, 경쟁, 관람, 커뮤니티, 플랫폼 경험이 결합된 복합적 현상으로 이해된다[21].

최근 문헌 검토 연구에서는 e스포츠 연구가 게임 연구, 스포츠 경영, 커뮤니케이션, 심리학, 교육, HCI 등 다양한 분야로 확장되고 있음을 보여준다. 또한 e스포츠가 경쟁 스포츠인지 여가 활동인지에 대한 논쟁 속에서 제도화, 규칙성, 신체성, 관중성 등이 주요 쟁점으로 다루어졌다[22]. 따라서 e스포츠 시장 동향을 분석할 때에는 게임의 인기뿐만 아니라 리그 구조, 조직화, 글로벌 제도, 참여 생태계를 함께 살펴볼 필요가 있다.

국내의 산업 동향 분석에서는 특허, 뉴스, 정책자료, 기관 보도자료와 같은 현장 기반 자료가 중요하다. KIPRIS는 기술 개발 방향을 파악하는 데 적합하며[23], KeSPA 보도 자료는 국내 e스포츠 운영 동향을 확인하는 데 활용될 수 있다. 또한 Global Esports Federation 자료는 국제 대회와 글로벌 협력, 제도화 흐름을 분석하는 데 유용하다.

토픽 모델링을 활용한 산업 동향 연구는 비정형 문서에서 잠재 주제를 도출한다는 점에서 본 연구와 관련된다. 기존 연구는 KIPRIS 특허 데이터를 대상으로 TF-IDF와 LDA를 적용해 NFT 산업의 핵심 주제를 분석한 반면[24], 본 연구는 이를 e스포츠 분야에 적용하되, 분석 대상을 특허, 보도자료, 글로벌 뉴스로 확장하여 산업 동향을 보다 입체적으로 파악하고자 한다.

2. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF는 문서 집합에서 특정 단어의 중요도를 산출하기 위한 대표적인 정보검색 가중치 기법이다. 단어 빈도(TF)는 특정 단어가 문서 내에서 얼마나 자주 등장하는지를 의미하고, 역문서 빈도(IDF)는 해당 단어가 전체 문서에서 얼마나 희소한지를 나타낸다. 따라서 여러 문서에서 지나치게 일반적으로 등장하는 단어의 가중치는 낮아지고, 특정 문서나 문서군의 특징을 설명하는 단어의 가중치는 높아진다.

본 연구에서는 KIPRIS 특허 자료, KeSPA 보도자료 및 Global Esports 뉴스 자료를 자료원별로 전처리하였다. Global Esports의 영문 어휘는 영한 e스포츠 도메인 사전을 활용하여 대응하는 한국어 개념으로 번역 및 매핑한 후, 자료원별 TF-IDF 값을 산출하였다. 이후 문서 수와 점수 규모의 차이에 따른 편향을 완화하기 위해 자료원 내부에서 TF-IDF 점수를 정규화하고, 세 자료원에 각각 동일한 1/3의 가중치를 적용하여 통합 핵심 키워드를 도출하였다. 이를 통해 특정 자료원에 편중되지 않고 각 자료

원의 주요 키워드를 균형 있게 반영하였다.

3. LDA(Latent Dirichlet Allocation)

LDA는 문서가 여러 잠재 토픽의 혼합으로 구성되어 있고, 각 토픽은 단어의 확률분포로 표현된다고 가정하는 생성 확률 모델이다. LDA는 문서-단어 행렬을 입력으로 하여 각 문서의 토픽 분포와 각 토픽의 단어 분포를 추정하며, 대규모 문서 집합에서 사람이 직접 분류하기 어려운 주제 구조를 발견하는 데 활용된다.

LDA는 pLSA가 가진 과적합 문제를 완화하기 위해 디리클레 사전분포를 도입한 모델로 설명될 수 있다[25]. 이로써 제한된 문서 수에서도 비교적 안정적인 토픽 분포 추정이 가능하며, 산업 동향 분석, 특허 분석, 뉴스 분석, 정책 의제 분석 등 다양한 텍스트 마이닝 연구에 활용되어 왔다. 본 연구에서는 통합 키워드 데이터에 기반하여 3개의 토픽을 도출하고, 각 토픽의 주요 키워드와 산업적 의미를 해석하였다.

III. The Proposed Scheme

1. Research Design

전체 연구 절차는 [Table 1]의 절차와 같이 데이터 수집, 전처리, TF-IDF 분석, 핵심 키워드 도출, LDA 토픽 모델링, 결과 해석의 여섯 단계로 구성된다. 먼저 KIPRIS, KeSPA, Global Esports에서 e스포츠 산업 관련 텍스트 자료를 수집하였다. 이후 Python 기반 전처리 코드를 활용하여 형태소 분석, 불용어 제거, 사용자 정의어 보존을 수행하고, Global Esports의 영문 어휘를 한국어 대응 개념으로 번역 및 매핑하여 단일 언어 특징 공간을 구축하였다. 전처리한 데이터를 통합 알고리즘을 통해 TF-IDF 값을 산출하고, 자료원 내부 정규화와 동일 가중치를 적용하여 통합 핵심 키워드를 도출하였다. LDA 토픽 모델링에서는 총 500개의 원문 문서를 분석 단위로 활용하고, 토픽 일관성, 혼잡도 및 안정성 등의 지표를 비교하여 최적 토픽 수를 결정하였다. 마지막으로 토픽별 주요 키워드와 자료원별 특성을 종합하여 e스포츠 산업에서 나타나는 주요 담론과 구조적 특징을 해석하였다.

Table 1. Research Procedure

Step	Procedure	Output
1	Web crawling	Raw text data
2	Preprocessing	Cleaned corpus
3	TF-IDF analysis	Top 50 keywords by source
4	LDA topic modeling	Three latent topics
5	Interpretation	Industry trend implications

2. Data Collection

자료원은 e스포츠 시장의 기술, 국내 제도권, 글로벌 동향을 균형 있게 반영할 수 있도록 선정하였다. 선정 결과 e스포츠 시장 동향을 다각적으로 분석하기 위해 KIPRIS 특허 자료, KeSPA 보도자료, Global Esports 뉴스 자료를 수집하였다. KIPRIS 자료는 e스포츠 관련 기술과 서비스 구현 방향을 확인하기 위한 목적으로 사용되었고, KeSPA 자료는 국내 대회 운영, 선수, 협회, 아카데미, 국가대표 체계와 관련한 제도적 흐름을 확인하기 위해 활용되었다. Global Esports 뉴스 자료는 국제 대회, 글로벌 거버넌스, 회원국, 파트너십, 지역별 행사와 같은 국제적 흐름을 파악하기 위한 자료원으로 선정하였다. 수집 자료는 시기별 키워드 또는 토픽 변화 분석보다는 수집자료에서 관찰된 주요 주제 수준으로 제한한다.

3. Web Crawling

KIPRIS 자료는 지식재산정보 검색 서비스에서 e스포츠 관련 특허를 검색한 뒤 특허 제목, 날짜 및 본문 정보를 수집하였다. KeSPA와 Global Esports 자료는 각 웹사이트의 보도자료 및 뉴스 게시글을 대상으로 크롤링을 수행하여 날짜, 제목 및 본문 정보를 정리하였다. 크롤링된 데이터는 엑셀 파일로 저장한 후 Python 기반 텍스트 분석 절차에 투입하였다.

각 자료원은 KIPRIS의 기술·서비스 개발 동향, KeSPA의 국내 리그 운영 및 인재 양성, Global Esports의 국제 협력과 글로벌 거버넌스 등 서로 다른 산업적 관점을 반영한다. 수집 후에는 제목과 본문이 동일한 중복 문서를 제거하고, 제목과 본문을 결합하여 각 게시물을 하나의 독립된 분석 문서로 구성하였다. 수집된 데이터는 [Table 2]와 같이 KIPRIS 40건, KeSPA 289건, Global Esports 171건으로 총 500건이며, 전처리 후 TF-IDF와 LDA 토픽 모델링에 활용하였다.

Table 2. Data Collection sources

Source	No. of documents	Purpose
KIPRIS	40	To identify eSports-related technologies, systems, and service models
KeSPA	289	Domestic leagues, institutions, players, and education programs
Global Esports	171	International events, federation governance, and global cooperation
Total	500	Integrated text-mining dataset for eSports Industry trend analysis

4. Preprocessing

웹 크롤링 후 전처리 단계에서는 제목과 본문이 동일한 중복 문서를 제거하고, 결측값을 빈 문자열로 대체한 뒤 제목과 본문을 결합하여 분석 문서를 구성하였다. 이후 영문 소문자 변환, URL 및 특수문자 및 웹페이지 메타데이터 제거, 복합어 보존과 불용어 제거를 수행하였다. 국내 자료인 KIPRIS와 KeSPA는 정규표현식을 이용하여 한글 및 영문 토큰을 추출하고, 복수 표현과 조사, 일부 서술형 표현을 규칙에 따라 제거하였다. Global Esports 자료는 영문 단어 복합어를 추출한 뒤 소유격과 일부 복수형을 정리하였다. 또한 KeSPA의 기관 보도자료 표현과 KIPRIS의 특허 문서 표현을 자료원별 불용어로 지정하여 제거하였다.

영문 자료와 한국어 자료를 동일한 언어 공간에서 분석하기 위해 연구자 검수 기반의 영한 e스포츠 도메인 사전을 구성하고, 추출된 영문 어휘를 e스포츠 분야에서 상응하는 한국어 대표 개념으로 번역 및 매핑하였다. 동의어, 복수형 및 파생형은 사전에 정의한 대표 개념으로 통합하여 최종 분석 어휘를 한국어 중심의 단일 개념 체계로 통일하였다. game, games, gaming은 ‘게임’으로, player, players, athlete, athletes는 ‘선수’로, tournament, championship, competition은 ‘대회’로, event, events는 ‘이벤트’로 변환하였다. 이를 통해 한국어와 영어로 표현된 동일한 개념이 서로 다른 키워드로 계산되는 문제를 방지하였다. 매핑 사전에는 영문 원어, 한국어 통합 개념, 사용 여부, 처리 유형 및 제외 사유를 기록하고, 모든 문서에 동일한 변환 기준을 적용하여 번역 및 개념 통합 과정의 일관성과 재현성을 확보하였다.

또한 동일한 의미를 가진 표기 변형이 서로 다른 키워드로 계산되는 문제를 방지하기 위해 정규표현식을 이용한 표준화 과정을 수행하였다. 먼저 모든 영문을 소문자로 변환한 후, ‘e스포츠’, ‘E스포츠’, ‘e-스포츠’, ‘e 스포츠’, esports, e-sports, e sports를 모두 표준형인 ‘이스포츠’로 통일하였다. 분석 코드에서는 영문 표기에

\be[s-]?sports\b, 한영 혼합 표기에 e[s-]?스포츠, 한국어 표기에 이[s-]?스포츠 패턴을 사전에 정의하고, 각 패턴에 해당하는 표현을 re.sub() 함수를 이용하여 '이스포츠'로 치환하였다. 이후 표준화된 '이스포츠'는 연구 주제 자체를 나타내어 키워드 간 변별력이 낮으므로 불용어로 지정하여 분석에서 제외하였다.

'리그 오브 레전드', '이터널 리턴', '스트리트 파이터', '가상 현실', '빅 데이터', '국가 대표' 등 여러 어절로 구성되지만 하나의 의미 단위로 해석되어야 하는 표현은 사용자 정의 복합어 치환 규칙을 통해 단일 어휘로 보존하였다. 반면 Global Esports Games, Global Esports Tour, Commonwealth Esports Championships와 같은 특정 대회명은 일반 단어로 분리되기 전에 하나의 고유명사로 인식한 후 통합 산업 키워드 분석에서 제외하였다. 이를 통해 특정 행사명이 global, esports, games 등의 일반 어휘로 분리되어 '글로벌'과 '게임'의 중요도를 과도하게 높이는 문제를 방지하였다.

전처리가 완료된 문서는 TF-IDF 분석과 LDA 토픽 모델링에 각각 독립적으로 활용하였다.

5. TF-IDF analysis

키워드 선정 과정에서는 Global Esports와 KeSPA는 최소 2개 이상의 문서에 출현한 어휘를 사용하였으며, 문서 수가 상대적으로 적은 KIPRIS는 최소 1개 이상의 문서에 출현한 어휘를 사용하였다. 이후 선정된 후보 어휘에 영한 도메인 사전과 자료원별 대표어 변환 규칙을 적용하여 동의어와 표기 변형을 동일한 대표 개념으로 통합하였다. 각 원문에서 후보 어휘에 해당하는 단어만을 대표 개념으로 변환하여 문서별 출현 구조를 재구성하고, 이를 기반으로 자료원별 TF-IDF를 다시 산출하였다. 각 키워드의 자료원별 중요도는 해당 자료원에 포함된 전체 문서의 TF-IDF 평균값으로 계산하였다.

본 연구에서 수집한 문서는 Global Esports 171건, KeSPA 289건, KIPRIS 40건으로 자료원별 문서 수에 차이가 있었다. 이러한 차이로 인해 문서 수가 많은 자료원의 키워드가 통합 결과를 과도하게 지배하는 문제를 완화하기 위해, 자료원별 평균 TF-IDF 점수의 합이 1이 되도록 L1 정규화를 수행하였다. 이후 동일한 대표 개념을 기준으로 세 자료원의 정규화 점수를 결합하였으며, 특정 자료원에 출현하지 않은 키워드의 점수는 0으로 처리하였다. 최종 통합 TF-IDF 점수는 KIPRIS, KeSPA, Global Esports의 정규화 점수에 각각 동일한 1/3의 가중치를 적용하여 산출하였다. 이를 통해 자료원별 문서 수와 점수

규모의 차이가 통합 결과에 미치는 영향을 완화하고, 세 자료원의 핵심 키워드가 균형 있게 반영되도록 하였다. 마지막으로 통합 TF-IDF 점수를 내림차순으로 정렬하여 상위 50개 어휘를 최종 핵심 키워드로 선정하였다.

6. LDA-based Topic Modeling

LDA 기반 토픽 모델링은 500개 원문 문서를 분석 단위로 재구성하고 토픽 수를 coherence 등의 객관적 기준으로 결정하여 수행하였다. [Table 3]에 따라 본 연구에서는 토픽 수를 3개로 설정한 이유는 다음과 같다. 토픽 수 선정을 위해 K를 2에서 10까지 변화시키며 토픽 일관성, 퍼플렉시티, 토픽 다양성 및 토픽 안정성을 비교하였다. K=3은 평균이 가장 높은 모형은 아니었으나, 표준편차가 0.01로 가장 낮고 토픽 안정성이 0.9071로 가장 높게 나타났다. 또한 토픽 다양성이 0.8933으로 가장 높아 토픽 간 구분성이 우수하였다. 퍼플렉시티 역시 K=2에서 K=3으로 증가할 때 크게 감소한 이후 개선 폭이 제한적으로 나타났다. 이에 따라 본 연구에서는 결과의 일관성, 재현성, 토픽 간 구분성 및 모형의 간결성을 종합적으로 고려하여 최적 토픽 수를 K=3으로 선정하였다. 500개 원문 문서를 분석 단위로 재구성하고 토픽 수를 coherence 등의 객관적 기준으로 결정하여 수행하였다.

Table 3. LDA Topic Number Evaluation

K	Mean	SD	Perplexity	Topic Diversity	Topic Stability
2	0.36	0.06	60.3825	0.8533	0.7696
3	0.41	0.01	48.3034	0.8933	0.9071
4	0.41	0.02	46.8816	0.8467	0.6305
5	0.43	0.03	46.7247	0.8747	0.4405
6	0.44	0.04	48.6824	0.8578	0.4052
7	0.42	0.03	51.0399	0.8514	0.3335
8	0.44	0.04	52.0146	0.8650	0.3187
9	0.45	0.07	52.3759	0.8563	0.3191
10	0.44	0.03	53.8435	0.8267	0.2818

토픽 다양성은 서로 다른 토픽의 상위 키워드가 얼마나 중복되지 않는지를 나타내는 것으로서, 값이 1에 가까울수록 토픽 간 키워드 중복이 적고 주제 구분성이 높음을 의미한다. 토픽 수가 K 이고, j 번째 토픽의 상위 M 개 키워드 집합을 T_j 라고 하면 다음과 같은 계산식을 제시한다 (1).

$$D_K = \frac{\left| \bigcup_{j=1}^K T_j \right|}{K \times M} \quad (1)$$

토픽 안정성은 랜덤 시드를 다르게 하여 LDA를 반복 실행했을 때, 유사한 토픽이 다시 나타나는지를 측정함으로써, 이때 실행마다 토픽의 배열 순서가 달라질 수 있으므로 헵가리안 알고리즘을 이용하여 토픽 간 최적 대응 관계를 설정한 후, 대응된 토픽들의 평균 유사도를 산출하였다. 최종 안정성은 5회 반복 실행에서 생성된 10개 실행 조합의 평균값으로 계산하였다. 먼저 두 번의 실행 r 과 s 에서 도출된 토픽의 상위 키워드 집합 간 자카드 유사도를 계산한다 (2).

$$J(T_i^{(r)}, T_j^{(s)}) = \frac{|T_i^{(r)} \cap T_j^{(s)}|}{|T_i^{(r)} \cup T_j^{(s)}|} \quad (2)$$

IV. Analysis Results

1. TF-IDF Keyword Analysis Result

자료원별 TF-IDF 상위 10개 키워드는 [Table 4]와 같다. KIPRIS에서는 ‘게임’, ‘경기’, ‘영상’, ‘베팅’, ‘데이터’, ‘정보’, ‘선수’, ‘중계’, ‘서버’, ‘실시간’이 상위에 포함되었다. 특히 영상, 데이터, 정보, 중계, 서버 및 실시간과 같은 기술 관련 키워드가 나타나 특히 자료에서는 경기 콘텐츠 뿐만 아니라 데이터 처리, 영상 제공, 실시간 중계 및 서비스 구현과 관련된 기술적 담론이 강조되고 있음을 확인할 수 있다. KeSPA에서는 ‘대회’, ‘LCK’, ‘선수’, ‘팀’, ‘경기’, ‘종목’, ‘아카데미’, ‘온라인’, ‘교육’, ‘결승’이 주요 키워드로 도출되었다. 이는 국내 협회 보도자료가 리그와 대회 운영, 선수·팀 관리 및 아카데미를 통한 인재 양성에 중점을 두고 있음을 나타낸다. Global Esports에서는 ‘글로벌’, ‘세계’, ‘게임’, ‘커뮤니티’, ‘이벤트’, ‘선수’, ‘대회’, ‘팀’, ‘파트너십’, ‘청소년’이 상위에 나타났다. 이는 해당 자료에서 국제 협력, 글로벌 커뮤니티, 대회 운영 및 청소년 참여와 관련된 담론이 주로 관찰되었음을 보여준다. 이러한 결과는 세 자료원이 각각 글로벌 협력, 국내 리그 운영, 기술·데이터 서비스라는 서로 다른 산업적 관점을 반영하고 있음을 보여준다.

Table 4. Top TF-IDF Keywords by Data Source

Rank	KIPRIS	KeSPA	Global Esports
1	게임	대회	글로벌
2	경기	lck	세계
3	영상	선수	게임
4	베팅	팀	커뮤니티
5	데이터	경기	이벤트
6	정보	종목	선수
7	선수	아카데미	대회
8	중계	온라인	팀
9	서버	교육	파트너십
10	실시간	결승	청소년

2. Integrated Keyword Analysis Results

세 자료원의 TF-IDF 점수를 자료원별로 정규화하고 동일한 가중치로 결합하여 통합 핵심 키워드를 도출하였다. [Table 5]에서 통합 TF-IDF 상위 15개를 살펴보면 ‘게임’이 0.017798로 가장 높았으며, ‘경기’ 0.012999, ‘선수’ 0.012982, ‘이벤트’ 0.009740, ‘대회’ 0.009589 순으로 나타났다. 이어 ‘글로벌’, ‘팀’, ‘영상’, ‘콘텐츠’, ‘실시간’, ‘교육’, ‘베팅’, ‘중계’, ‘데이터’, ‘온라인’이 포함되었다. 이는 수집 자료에서 e스포츠의 경기, 선수, 대회 중심 경쟁 구조와 함께 글로벌 확장, 디지털 콘텐츠, 실시간 중계 및 데이터 기반 서비스가 주요 담론으로 나타났음을 보여준다.

문서 빈도를 살펴보면 ‘대회’가 314건으로 가장 많은 문서에서 확인되었으며, ‘선수’ 302건, ‘팀’ 282건, ‘글로벌’ 251건, ‘게임’ 234건, ‘경기’ 216건 순으로 나타났다. 반면 ‘영상’ 29건, ‘데이터’ 25건, ‘베팅’ 4건 등 기술 및 서비스 관련 키워드는 문서 빈도가 상대적으로 낮았지만 통합 TF-IDF 상위권에 포함되었다. 이는 해당 키워드가 전체 자료에서 광범위하게 등장하기보다는 특정 등 특정 자료원에서 상대적으로 높은 중요도를 가진다는 것을 의미한다. 따라서 분석 자료에 나타난 e스포츠 산업 담론은 경기, 대회 운영을 중심으로 형성되어 있으면서도, 영상, 콘텐츠, 실시간 중계, 데이터와 같은 기술 기반 서비스가 산업 확장의 주요 요소로 함께 나타난 것으로 해석할 수 있다. 종합하면, e스포츠 산업 담론은 경기, 대회, 선수 등 경쟁 및 운영 구조를 중심으로 형성되어 있다. 동시에 영상, 데이터, 실시간 중계 등은 문서 빈도는 낮지만, 특정 자료원에서 높은 중요도를 보여 기술 기반 서비스가 산업 확장의 주요 축임을 시사한다.

Table 5. Integrated TF-IDF Top Keywords

Rank	Keyword	Integrated TF-IDF	Document Frequency
1	게임	0.017798	234
2	경기	0.012999	216
3	선수	0.012982	302
4	이벤트	0.009740	183
5	대회	0.009589	314
6	글로벌	0.008197	251
7	팀	0.008093	282
8	영상	0.007545	29
9	콘텐츠	0.006907	75
10	실시간	0.006665	70
11	교육	0.006535	92
12	베팅	0.006022	4
13	중계	0.005918	44
14	데이터	0.005886	25
15	온라인	0.005792	163

[Fig. 1] 내용과 같이 자료원별 분기 수집 현황을 살펴보면, 2020년부터 2022년까지는 Global Esports 자료가 주로 나타났으며, 2023년 4분기 이후부터 KeSPA 자료의 비중이 크게 증가하는 형태를 보였다. 특히 2024년과 2025년 구간에서는 KeSPA 자료가 높은 빈도로 나타났는데, 이는 해당 기간 국내 리그 운영, 대회 개최, 선수 관리 및 협회 활동과 관련된 보도자료가 집중적으로 생산되었음을 보여준다. 이에 반해 KIPRIS 자료는 전체 자료에서 상대적으로 낮은 비중을 보였으나, 특히 자료의 특성상 게임, 경기, 영상, 데이터, 정보, 중계 및 서버 등 기술·서비스 구현과 관련된 내용을 확인할 수 있었다. Global Esports 자료는 전체 기간에 걸쳐 비교적 지속적으로 분포하였으며, 일부 분기에서 수집량이 증가해 국제 대회와 글로벌 e스포츠 관련 뉴스가 특정 시기에 집중되었음을 보여준다.

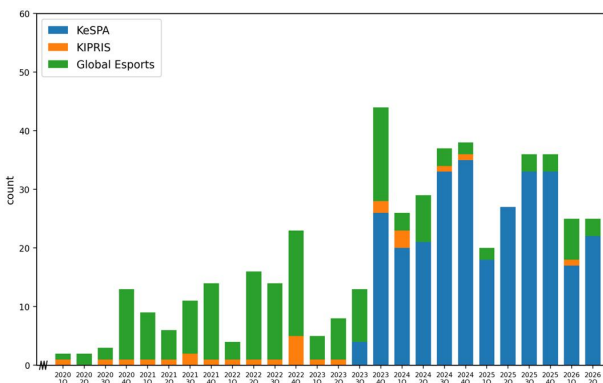


Fig. 1. Quarterly Distribution of Collected Data by Source

[Fig. 2]의 내용과 같이 자료원별 평균 TF-IDF를 L1 정규화하고 동일한 가중치를 적용하여 산출한 통합 TF-IDF 값에서는 ‘게임’이 0.017798로 가장 높게 나타났다. 그다음으로 ‘경기’ 0.012999, ‘선수’ 0.012982, ‘이벤트’ 0.009740, ‘대회’ 0.009589 순으로 나타나 게임과 경기, 선수 및 대회 운영과 관련된 개념이 수집자료의 핵심 키워드로 확인되었다. 또한 ‘글로벌’, ‘팀’, ‘영상’, ‘콘텐츠’, ‘실시간’이 상위 10개 키워드에 포함되었으며, ‘교육’, ‘베팅’, ‘중계’, ‘데이터’, ‘온라인’, ‘정보’, ‘서버’, ‘모바일’ 등의 키워드도 상위권에 나타났다. 이는 수집자료에서 관찰된 e스포츠 산업의 주요 담론이 경기와 선수 중심의 경쟁 구조뿐만 아니라 글로벌 확장, 영상 콘텐츠와 실시간 중계, 데이터·정보 처리 및 온라인·모바일 서비스 영역까지 폭넓게 연결되어 있음을 보여준다. 여기서 통합 TF-IDF 값은 단순한 출현 빈도가 아니라 세 자료원의 정규화 점수에 동일 가중치를 적용한 상대적 중요도이므로, 문서 수가 많은 특정 자료원의 영향이 완화된 결과로 해석할 수 있다.

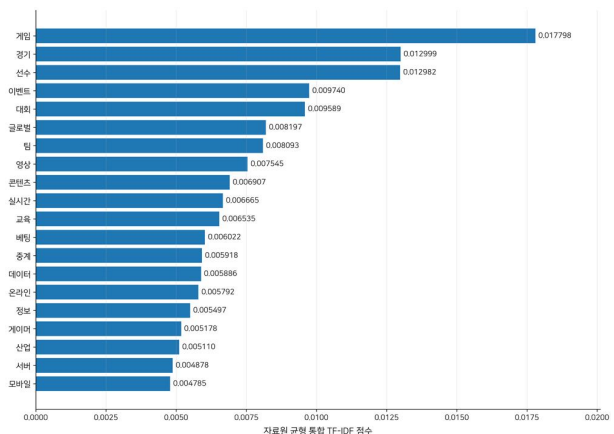


Fig. 2. Top 20 Keywords by Integrated TF-IDF Score

[Fig. 3]는 통합 TF-IDF 상위 50개 키워드의 통합점수를 단어 크기와 색상의 농도 차이로 시각화한 워드클라우드이다. ‘게임’이 가장 크게 나타났으며, ‘경기’와 ‘선수’가 그 뒤를 이었다. 또한 ‘이벤트’, ‘대회’, ‘글로벌’, ‘팀’, ‘영상’, ‘콘텐츠’, ‘실시간’ 등의 키워드가 주요하게 나타났고, ‘교육’, ‘베팅’, ‘중계’, ‘데이터’, ‘온라인’, ‘정보’ 등도 함께 확인되었다. 이는 통합 TF-IDF 분석 결과를 시각적으로 보여주는 것으로, 수집자료에서 관찰된 e스포츠 산업의 주요 담론이 게임·경기·선수를 중심으로 한 경쟁 구조뿐만 아니라 대회 운영, 글로벌 확장, 영상·실시간 콘텐츠, 데이터 기반 기술 서비스 및 교육 영역과도 연결되어 있음을 보여준다.

3. LDA Topic Modeling Analysis Results

첫 번째 토픽은 ‘게임’, ‘경기’, ‘데이터’, ‘영상’, ‘사용자’, ‘서버’, ‘콘텐츠’, ‘배팅’, ‘정보’, ‘실시간’, ‘중계’, ‘가상현실’, ‘예측’, ‘서비스’ 등의 키워드로 구성되어 ‘기술·데이터 서비스’로 명명하였다. 이 토픽은 경기 데이터와 영상 정보를 처리하고 사용자에게 실시간 콘텐츠와 중계 서비스를 제공하는 기술적 구조를 반영한다. 또한 서버, 가상현실, 예측, 인터페이스 및 단말 관련 키워드가 포함되어 있어 e스포츠가 데이터 처리와 디지털 플랫폼을 기반으로 다양한 기술 서비스와 결합되고 있음을 보여준다.

세 번째 토픽은 ‘게임’, ‘세계’, ‘글로벌’, ‘커뮤니티’, ‘대회’, ‘이벤트’, ‘선수’, ‘팀’, ‘국가’, ‘위원회’, ‘파트너십’, ‘회원’ 등의 키워드로 구성되어 ‘글로벌 협력·거버넌스’로 명명하였다. 이 토픽은 국제 대회와 이벤트를 중심으로 글

자료원별 평균 토픽 확률을 비교한 결과, Global Esports 문서는 ‘글로벌 협력·거버넌스’ 토픽의 평균 확률이 0.9989로 나타났으며, 171개 문서 모두에서 해당 토픽이 우세하였다. KeSPA 문서는 ‘국내 리그·인재 양성’ 토픽의 평균 확률이 0.9835로 나타났으며, 289개 문서 모두에서 해당 토픽이 우세하였다. KIPRIS 문서는 ‘기술·데이터 서비스’ 토픽의 평균 확률이 0.7879로 가장 높았고, 전체 40개 문서 가운데 31개 문서에서 해당 토픽이 우세한 것으로 나타났다. 나머지 9개 KIPRIS 문서에서는 ‘국내 리그·인재 양성’ 토픽이 우세하였다. 이러한 결과는 도출된 세 토픽이 각 자료원의 문서 특성을 뚜렷하게 반영하고 있음을 보여준다. 종합하면, LDA 분석을 통해 e스포츠 산업 담론이 기술·데이터 서비스, 국내 리그·인재 양성, 글로벌 협력·거버넌스의 세 영역으로 구분되며 각 자료원의 특성이 뚜렷하게 반영됨을 확인하였다.

Topic	Main keywords	Interpretation
기술·데이터 서비스	게임	A structure that processes match data and video and provides users with real-time content and technology-based services.
	경기	
	데이터	
	영상	
	사용자	
	서버	
	콘텐츠	
국내 리그·인재 양성	베팅	A talent development system encompassing domestic league and tournament operations, player management, and academy-based training.
	선수	
	대회	
	LCK	
	운영	
	경기	
	PC	
글로벌·협력 거버넌스	아카데미	A cooperation and governance structure centered on international tournaments and events and the global esports community.
	리그	
	게임	
	세계	
	글로벌	
	커뮤니티	
	대회	
	이벤트	
	선수	
	팀	

4. Discussion

분석 결과를 종합하면, 수집자료에서 관찰된 e스포츠 산업의 주요 담론은 경기 데이터와 영상 콘텐츠를 기반으로 한 ‘기술·데이터 서비스’, 국내 리그 운영과 교육을 연계한 ‘국내 리그·인재 양성’, 국제 대회와 커뮤니티를 중심으로 한 ‘글로벌 협력·거버넌스’의 세 영역으로 구분된다. 통합 TF-IDF 분석에서도 ‘게임’, ‘경기’, ‘선수’, ‘이벤트’, ‘대회’와 함께 ‘글로벌’, ‘영상’, ‘콘텐츠’, ‘실시간’, ‘교육’, ‘데이터’ 등의 키워드가 상위권에 나타났다. 이는 e스포츠가 경기와 대회 운영뿐만 아니라 기술 서비스, 교육 체계 및 국제적 협력이 상호 결합된 복합적인 산업 구조를 형성하고 있음을 보여준다. 또한 이러한 결과는 e스포츠가 전통 스포츠와 유사하게 리그, 협회, 국가대표 및 국제대회 체계를 갖추며 제도화되고 있다는 기존 연구의 논의와도 연결된다.

KIPRIS 특허 자료에서는 ‘게임’, ‘경기’, ‘영상’, ‘베팅’, ‘데이터’, ‘정보’, ‘중계’, ‘서버’, ‘실시간’ 등의 키워드가 주요하게 나타났다. LDA 분석에서도 ‘기술·데이터 서비스’ 토픽의 평균 확률이 0.7879로 가장 높았으며, 전체 40개 특허 문서 가운데 31개 문서에서 해당 토픽이 우세하였다. 이 토픽에는 데이터, 영상, 사용자, 서버, 콘텐츠, 베팅, 정보, 실시간, 중계, 가상현실, 예측 및 인터페이스 등의 키워드가 포함되었다. 이는 e스포츠 관련 특허가 경기 자체보다 경기 데이터의 처리와 분석, 영상 및 실시간 중계, 사용자 인터페이스, 서버 기반 콘텐츠 제공과 같은 기술적 구현에 중점을 두고 있음을 보여준다. 따라서 e스포츠는 경기 콘텐츠와 관람 경험을 기반으로 하면서도 데이터 처리, 디지털 플랫폼 및 정보 서비스가 결합된 기술 융합 산업으로 이해할 필요가 있다.

KeSPA 자료에서는 ‘대회’, ‘LCK’, ‘선수’, ‘팀’, ‘경기’, ‘종목’, ‘아카데미’, ‘교육’ 등의 키워드가 중요하게 나타났으며, LDA 분석에서도 ‘국내 리그·인재 양성’ 토픽의 평균 확률이 0.9835로 확인되었다. 이는 국내 e스포츠 담론이 리그와 대회의 운영, 선수 및 종목 관리, 아카데미와 교육을 통한 인재 육성을 중심으로 형성되어 있음을 보여준다. 한국e스포츠협회는 2026년 5월 18일 아이치·나고야 아시안게임에 출전할 e스포츠 국가대표 선수 및 지도자 최종 후보 명단을 공개하였으며, 공식 이의 신청과 대한체육회의 승인 절차를 거쳐 최종 엔트리를 확정할 예정이라고 밝혔다. 이는 국내 e스포츠가 민간 대회 중심의 경쟁 활동을 넘어 국가대표 선발, 선수 관리 및 국제 종합대회 참가 절차를 갖춘 제도권 스포츠로 확장되고 있음을 보여주는 사례이다. 또한 아카데미와 교육 관련 키워드는 국내 e스포

츠 생태계에서 전문 선수와 지도자를 육성하기 위한 교육 기반이 중요하게 다루어지고 있음을 나타내며, 장기적으로 프로 리그와 국가대표 단계로 이어지는 인재 성장 경로의 형성 가능성을 시사한다.

Global Esports 자료에서는 ‘글로벌’, ‘세계’, ‘게임’, ‘커뮤니티’, ‘이벤트’, ‘선수’, ‘대회’, ‘팀’, ‘파트너십’, ‘청소년’ 등의 키워드가 중요하게 나타났다. LDA 분석에서도 Global Esports의 171개 문서 모두에서 ‘글로벌 협력·거버넌스’ 토픽이 우세하였으며, 평균 토픽 확률은 0.9989로 확인되었다. 이는 해당 자료에서 국제 대회와 이벤트뿐만 아니라 국가, 회원 조직, 위원회, 파트너십 및 글로벌 커뮤니티의 연결이 중요하게 다루어지고 있음을 의미한다. 따라서 Global Esports 자료에서 관찰되는 e스포츠 담론은 개별 경기나 지역 대회의 운영을 넘어 국제 조직 간 협력, 글로벌 네트워크 형성 및 국제적 거버넌스 구축과 밀접하게 관련된 것으로 해석할 수 있다.

방법론적으로 본 연구는 특허, 협회 보도자료 및 글로벌 뉴스라는 성격이 서로 다른 자료원을 함께 분석함으로써 단일 자료원만으로는 파악하기 어려운 기술적, 제도적, 국제적 담론을 동시에 확인하였다는 의의가 있다. 특히 자료원별 TF-IDF 점수를 L1 정규화하고 동일한 가중치를 적용함으로써 문서 수가 많은 KeSPA 자료가 통합 결과를 과도하게 지배하는 문제를 완화하였다. 또한 LDA 분석에서는 500개의 원문 문서를 개별 분석 단위로 사용하고 문서 길이 표준화와 자료원별 역비를 가중치를 적용하여 자료원 간 규모 차이를 통제하였다.

V. Conclusions

본 연구는 KIPRIS 특허 40건, KeSPA 보도자료 289건, Global Esports 뉴스 171건 등 총 500개의 문서를 대상으로 TF-IDF와 LDA 토픽 모델링을 적용하여 e스포츠 산업의 주요 담론을 분석하였다. 자료원별 TF-IDF 분석에서는 KIPRIS의 기술적 관점, KeSPA의 국내 리그 운영 및 인재 양성 관점, Global Esports의 국제 협력 및 거버넌스 관점이 구분되어 나타났다. 자료원별 평균 TF-IDF를 L1 정규화하고 동일한 가중치로 결합한 결과, ‘게임’, ‘경기’, ‘선수’, ‘이벤트’, ‘대회’가 가장 중요한 통합 키워드로 나타났으며, ‘글로벌’, ‘팀’, ‘영상’, ‘콘텐츠’, ‘실시간’, ‘교육’, ‘데이터’ 등도 주요 키워드로 확인되었다. LDA 분석에서는 ‘기술·데이터 서비스’, ‘국내 리그·인재 양성’, ‘글로벌 협력·거버넌스’의 세 가지 토픽이 도출되었다. 이는 수

집자료에서 관찰된 e스포츠 산업의 주요 담론이 데이터 처리와 실시간 콘텐츠를 기반으로 한 기술 서비스, 리그·대회 운영과 교육을 연계한 국내 인재 양성, 국제 대회와 커뮤니티를 중심으로 한 글로벌 협력 구조로 구성되어 있음을 보여준다. 따라서 e스포츠는 경기 콘텐츠를 중심으로 기술, 교육, 제도 및 국제 협력 요소가 결합된 복합적인 산업으로 이해할 필요가 있다.

다만 본 연구는 분석 대상이 특허, 협회 보도자료 및 글로벌 뉴스로 한정되어 있어 연구 결과를 e스포츠 산업 전체의 일반적 특성으로 확대하여 해석하기 어렵다. 특히 자료원별 우세 토픽이 KIPRIS의 '기술·데이터 서비스', KeSPA의 '국내 리그·인재 양성', Global Esports의 '글로벌 협력·거버넌스'와 뚜렷하게 대응하였다는 점에서, 도출된 토픽에는 산업의 잠재적 주제뿐만 아니라 각 자료원의 작성 목적과 문서 특성도 반영되었을 가능성이 있다. 또한 TF-IDF와 LDA는 단어의 출현 빈도와 동시 출현 구조를 기반으로 하므로 문장의 구체적인 맥락, 의미 관계 및 시간에 따른 담론의 변화를 충분히 반영하지 못한다. 향후 연구에서는 게임사, 프로구단, 스트리밍 플랫폼, 소셜미디어 및 이용자 커뮤니티 등으로 자료원을 확대하고, 시계열 토픽 분석, 키워드 네트워크 분석, 감성 분석 및 BERTopic과 같은 임베딩 기반 토픽 모델링을 함께 적용하여 e스포츠 산업의 변화 구조를 더욱 정교하게 검증할 필요가 있다.

REFERENCES

- [1] J. Hamari and M. Sjöblom, "What is eSports and why do people watch it?," *Internet Research*, Vol. 27, No. 2, pp. 211-232, 2017. DOI: 10.1108/IntR-04-2016-0085
- [2] S. E. Jenny, R. D. Manning, M. C. Keiper, and T. W. Olrich, "Virtual(ly) Athletes: Where eSports Fit Within the Definition of Sport," *Quest*, Vol. 69, No. 1, pp. 1-18, 2017. DOI: 10.1080/00336297.2016.1144517
- [3] D. C. Funk, A. D. Pizzo, and B. J. Baker, "eSport Management: Embracing eSport Education and Research Opportunities," *Sport Management Review*, Vol. 21, No. 1, pp. 7-13, 2018. DOI: 10.1016/j.smr.2017.07.008
- [4] J. G. Reitman, M. J. Anderson-Coto, M. Wu, J. S. Lee, and C. Steinkuehler, "Esports Research: A Literature Review," *Games and Culture*, Vol. 15, No. 1, pp. 32-50, 2020. DOI: 10.1177/1555412019840892
- [5] T. L. Taylor, "Raising the Stakes: E-Sports and the Professionalization of Computer Gaming," Cambridge, MA: MIT Press, 2012. DOI: 10.7551/mitpress/8624.001.0001
- [6] Y. Seo and S. U. Jung, "Beyond Solitary Play in Computer Games: The Social Practices of eSports," *Journal of Consumer Culture*, Vol. 16, No. 3, pp. 635-655, 2016. DOI: 10.1177/1469540514553711
- [7] G. Freeman and D. Y. Wohn, "eSports as an Emerging Research Context at CHI: Diverse Perspectives on Definitions," *Proceedings of CHI EA 2017*, pp. 1601-1608, 2017. DOI: 10.1145/3027063.3053158
- [8] K. Hallmann and T. Giel, "eSports - Competitive Sports or Recreational Activity?," *Sport Management Review*, Vol. 21, No. 1, pp. 14-20, 2018. DOI: 10.1016/j.smr.2017.07.011
- [9] Olympics.com, "Esports Make Historic Medal Debut at 19th Asian Games in Hangzhou" 2023. <https://www.olympics.com/>
- [10] Olympic Council of Asia, "OCA Announces Titles for Hangzhou Asian Games Esports Competition," 2021. <https://oca.asia/>
- [11] Global Esports Federation, "Member Federations and Global Esports News," <https://www.globalesports.org/news/>
- [12] Korea e-Sports Association, "About KeSPA," <https://www.e-sports.or.kr/>
- [13] G. Pözlbauer and E. Auer, "Applied Patent Mining with Topic Models and Meta-data: A Comprehensive Case Study," *World Patent Information*, Vol. 67, Article 102065, 2021. DOI: 10.1016/j.wpi.2021.102065
- [14] P. Oldham, "The WIPO Patent Analytics Handbook: Chapter 7 Text Mining," *World Intellectual Property Organization*, 2022.
- [15] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," Cambridge University Press, 2008.
- [16] G. Salton and C. Buckley, "Term-weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, Vol. 24, No. 5, pp. 513-523, 1988. DOI: 10.1016/0306-4573(88)90021-0
- [17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.
- [18] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey," *Multimedia Tools and Applications*, Vol. 78, pp. 15169-15211, 2019. DOI: 10.1007/s11042-018-6894-4
- [19] M. Grootendorst, "BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure," *arXiv preprint arXiv:2203.05794*, 2022. DOI: 10.48550/arXiv.2203.05794
- [20] A. D. Pizzo, B. J. Baker, S. Na, M. A. Lee, D. Kim, and D. C. Funk, "eSport vs. Sport: A Comparison of Spectator Motives," *Sport Marketing Quarterly*, Vol. 27, No. 2, pp. 108-123, 2018. DOI: 10.32731/SMQ.272.062018.04
- [21] E. E. Cranmer, D. Han, M. Van Gisbergen, T. Jung, "Esports matrix: Structuring the esports research agenda. Computers in

- Human Behavior," Vol. 117, No. 106671, 2021. DOI: 10.1016/j.chb.2020.106671
- [22] J. Formosa, N. O'Donnell, E. M. Horton, S. Türkay, R. L. Mandryk, M. Hawks, and D. Johnson, "Definitions of Esports: A Systematic Review and Thematic Analysis," *Proceedings of the ACM on Human-Computer Interaction*, Vol. 6, CHI PLAY, pp. 1-45, 2022. DOI: 10.1145/3549490
- [23] Korea Intellectual Property Rights Information Service, "KIPRIS Introduction," <https://www.kipris.or.kr/>
- [24] S. N. Choi and W. Kim, "NFT(Non-Fungible Token) Patent Trend Analysis using Topic Modeling," *Journal of The Korea Society of Computer and Information*, Vol. 28, No. 12, pp. 41-48, 2023. DOI: 10.9708/jksci.2023.28.12.041
- [25] T. Hofmann, "Probabilistic Latent Semantic Indexing," *Proceedings of the 22nd Annual International ACM SIGIR Conference*, pp. 50-57, 1999. DOI: 10.1145/312624.312649

Authors



Tae-Min Lee is currently pursuing a bachelor's degree with a double major in Digital Twin Software and Esports at Dongyang University, Dongducheon, Republic of Korea. His research interests include

artificial intelligence, machine learning, esports analytics, and text mining.



Jeon-Young Kong received his Ph.D. degree in Cultural Contents from Konkuk University, Seoul, Republic of Korea. He is currently a professor in the Department of eSports at Dongyang University. His research interests

include game policy, esports education, esports marketing and business, cultural contents, and research on the effects of games and esports.



Woong Kim received his M.S. degree in Industrial Management Engineering from Korea University, Seoul, Republic of Korea, in 2016, and subsequently completed his doctoral coursework in the same field at

Korea University. He is currently a professor in the Department of Digital Twin Software at Dongyang University. His research interests include artificial intelligence, machine learning, logistics and transportation information systems, and supply chain management (SCM).