한국특수체육학회지 2022. 제30권 제2호. pp. 17-28 ISSN 1226-5063 (Print)
Korean Journal of Adapted Physical Activity, 2022. Vol. 30, No. 2, pp. 17-28. ISSN 2733-9831 (Online)

텍스트 마이닝 기반 토픽 모델링을 활용한 한국특수체육학회지 연구 동향 분석

배건환(한국체육대학교/박사과정) · **노형규***(한국체육대학교/교수)

초 록

본 연구는 지난 29년간 한국특수체육학회지에 게재된 연구들의 주제를 파악하고 시기별 이슈, 주제 출현 추이 및 특징을 분석하여 한국특수체육학회지의 전반적인 연구 동향을 파악하고자 하였다. 이를 위해 1993년부터 2021년까지 한국특수체육학회지에 게재된 951개 논문 중 영문초록이 존재하는 924개 논문을 대상으로 텍스트 마이닝 기반 토픽 모델링 분석을 실시하였다. Python 3.7 Package를 활용해 LDA(Latent Dirichlet Allocation) 알고리즘 기반 토픽 모델링 기법을 적용하였으며 분석 결과 '운동 프로그램 효과', '선수 사회적 지원', '사회·심리학적 접근', '특수학교(급) 현장', '선수 트레이닝', '아동과 부모'라는 총 6개의 토픽이 도출되었다. 또한 시계열에 따른 연구 동향을 파악하기 위해 시기를 6개 구간으로 구분하여 분석을 진행하였다. 시기별 토픽 점유율의 변화 추이를 살펴보면 '운동 프로그램 효과' 토픽은 제1시기부터 제3시기까지 일정 점유율을 유지(22.4%, 20.1%, 24.6%)하다 제4시기(14.5%) 에서 대폭 감소 후 제5시기(16.2%) 소폭 증가, 가장 근래인 제6시기(7%)에서는 대폭 감소하는 경향을 나타냈다. '선수 사회적 지원' 토픽은 제1시기(21.1%)에서 제2시기(9%)로 가면서 대폭 감소한 이후 제3,4시기(13.2%, 21.4%)를 거치며 다시 증가하였고 제5시기(21.4%) 정체 구간 이후 제6시기(22.2%)에서 소폭 증가하였다. '사회·심리학적 접근'은 제1시기(2.7%)부터 제4시기(13.8%)까지 꾸준히 증가하다 제5시기(9.2%)에서 감소한 이후 제6시기(19%)에서 다시 크게 반등하는 추이를 보였다. '특수학교(급) 현장' 토픽은 전반적으로 증감을 반복하며 시기별 점유율 상위권을 형성하는 경향을 보였으며 특히 제2시기(32.1%)에서의 폭발적 증가, 전체 시기 통합점유율(24.5%) 1위 등의 특징이 있었다. '선수 트레이닝'은 제4시기(6.9%)를 제외하면 지속적으로 10% 초중반대의 점유율을 유지하며 꾸준히 출현하는 경향을 나타냈다. 마지막으로 '아동과 부모' 토픽은 제1시기부터 제3시기까지(19%, 21.6%, 21%)는 점유율 20% 내외를 유지하였고 제4시기부터 제6시기까지(16.6%, 16.8%, 15.8%)는 16% 내외의 점유율을 나타내어 꾸준히 중위권을 형성하는 추이를 보였다. 이러한 결과는 기존 한국특수체육학회지 내 연구물의 현황과 추세를 분석 및 정리하여 특수체육이 향후 나아갈 방향성을 파악하고 전망할 수 있다는 점에서 의의가 있다. 본 연구 결과는 특수체육 연구 영역 확장 및 발전을 위한 후속 연구의 기초자료로 활용할 수 있을 것으로 사료된다.

주요어 : 한국특수체육학회지, 연구주제, 연구동향 분석, 텍스트 마이닝, 토픽 모델링

* rhk0019@knsu.ac.kr

www.kci.go.kr

Ⅰ. 서론

연구 동향 분석은 기존 연구 업적의 확인 및 평가에 더해 해당 연구 분야가 나아갈 방향성을 파악하고 전망 할 수 있는 반성적 작업이다(김기룡, 김삼섭, 2016). 학문분야의 역사가 비교적 짧을수록 정체성과 타당성의 문제를 고려하여 향후 학문 발전 방향에 대한 정보를 확보하는 과정이 더욱 필요한데(Sheldon, 1991), 그 중에서도 연구 동향 분석을 통한 정보 확보는 학문 정체성 수립과 발전 측면에서 필수적인 절차라고 볼 수 있다(정경환, 원영신, 구송광, 2015).

한국특수체육학회는 1990년 창립된 이후 2000년 체육학 분야 내 최초 국가인정 학회지로 등록되었으며 2003년(제 14회)과 2017년(제 21회) 2회에 걸쳐 세계특수체육학회(IFAPA)를 개최(최승권, 2018)한데 이어 2020년 창립 30주년을 맞이하는 등 1993년 첫 학술지 발행 이후 29년간 총 951개의 연구물을 배출(2022년 3월 기준)하며 발전을 거듭해왔다. 기존 연구물의 현황 및 추세를 체계적으로 분석하고 정리하는 작업은 학문 발전을 위해 매우 중요한 작업(김동남, 홍금석, 고진환, 전정환, 2017)인만큼 특수체육학 분야에서도 이러한 연구 동향 분석의 시도는 지속적으로 이루어져왔다(이경옥, 하지연, 2000; 정복자, 홍양자, 2006; 이현수, 박수경, 박진우, 윤석민, 최승오, 2011; 정경환 등, 2015; 진주연, 윤지은, 김은혜, 김지영, 2021). 그러나 대부분의 연구들이 특정 주제 또는 특정 시기에 한정하여 소수 논문을 분석한 것에 그쳤으며 최초 발행 학술지 게재 논문부터 최근까지의 논문 전수를 확보하여 동향을 분석한 연구는 전무한 실정이었다. 이로 인해 학술지를 대표하거나 특수체육에 대한 총체적 가치를 내포하기에는 다소 제약이 있었다. 따라서 학술지 발행 30주년을 앞둔 현 시점에서 방대한 수의 기존 연구물을 종합하여 연구 동향을 체계적으로 분석한다면 향후 학문 발전 방향에 더욱 풍부한 논제를 제공하고 의미 있는 시사점을 제시할 수 있을 것으로 판단된다.

한편 연구 동향 분석과 같이 방대한 양의 정보를 다루는 경우 빅데이터(big data) 분석 방법을 활용하게 되는데, 이러한 빅데이터 분석을 통해서는 사회적 흐름과 시대별 이슈를 종합적으로 분석하고 융합 가치 창출 및 다학제적 연구 접근 가능성 확인 등이 가능하다(전채남, 서일원, 2013). 이러한 빅데이터 분석 방법에는 데이터 마이닝, 텍스트 마이닝, 소셜네트워크분석, 클러스터분석 등이 있다(윤홍근, 2013). 그 중에서도 많은 양의 비정형 데이터를 효율적이고 체계적으로 다루기 위한 분석기법으로 최근 텍스트 마이닝(text mining)기법이 널리 활용되고 있는 추세다(반현정, 강병남, 김학선, 2021). 텍스트 마이닝 기법은 다량의 연구물을 종합하여 동향을 파악하는데 소요되는 시간을 단축시켜주고 분석 자료의 신뢰도를 확보할 수 있는 간편하고 효과적인 방법이다(김기룡, 김삼섭, 2016). 이러한 텍스트 마이닝 기반의 토픽모델링(topic modeling)은 연구자 주관성을 최소화하고 방대한 정보의 효율적 경향 분석 및 단어 간 관계에 따른 유의미한 토픽 추출을 가능하게 해준다(김지은, 백순근, 2016). 토픽 모델링은 빅데이터 내에서 특정한 토픽을 추출해내는 알고리즘으로 단어 사이의 연관성을 파악하고(Blei, Ng, & Jordan, 2003) 시기별 연구 동향 및 주제를 객관적으로 살펴볼 수 있는 LDA 기반 확률적 모델이다(반현정 등, 2021).

따라서 본 연구에서는 텍스트 마이닝 기반 토픽모델링의 LDA 알고리즘을 활용하여 도출된 데이터를 통해 29년간의 한국특수체육학회지 연구 동향을 분석하고자 한다. 본 연구는 1993년부터 2021년까지 한국특수체육학회지에 게재된 연구물 전체의 영문초록(abstract)을 크롤링(crawling)하여 특수체육 분야 내 시기별 이슈와 주제 출현 추이, 특징 등을 파악하고 이를 통해 향후 특수체육 연구 영역의 확장 및 발전을 위한 기초자료를 제시하는데 그 목적이 있다. 이를 위해 본 연구에서는 다음과 같이 연구문제를 설정하였다.

www.kci.go.kr

1) 29년간의 한국특수체육학회지에서 도출되는 연구주제 및 핵심어는 무엇인가?
2) 한국특수체육학회지의 연구주제 동향은 시계열적으로 어떠한 변화양상을 보이는가?

Ⅱ. 연구 방법

1. 연구자료

본 연구에서는 한국특수체육학회지의 창간호인 1993년 1권 1호부터 2021년 29권 4호에 게재된 총 951개 논문의 영문초록을 분석 자료로 선정하였다. 웹 크롤링을 통해 수집 가능한 국문초록 데이터에 비해 영문초록 데이터가 약 350여개 이상 더 많았기에 영문초록을 분석 대상 자료로 결정하는 것이 더욱 폭넓은 연구 동향 분석에 용이할 것이라 판단하여 영문초록을 최종 분석 대상 자료로 선정하였다. 초록은 연구 전문의 훌륭한 대체물로 인정되며(문주영, 2009) 연구의 핵심 주제어들을 포함하는 중요한 부분이다(김경진, 2018). 따라서 초록은 논문의 전반적 특성을 효율적으로 파악할 수 있는 유용한 분석 대상 자료다(김기룡, 김삼섭, 2016).

분석의 주된 재료가 되는 데이터 수집은 웹 크롤링(Web Crawling)이라고 하는 문서 수집 기술을 통해 이루어졌으며 이를 통해 총 951개 논문 중 영문초록 데이터 924개를 수집하였다. 자료 수집은 한국학술지인용색인(Korea Citation Index)와 한국교육학술정보원의 학술연구정보서비스(Research Information Sharing Service)의 검색 엔진을 이용하였으며 영문초록이 존재하지 않거나 웹 사이트 내에 중복 등록된 논문의 경우 분석 대상에서 제외하였다.

2. 자료처리

웹 크롤링을 통해 수집된 텍스트 자료는 모두 엑셀(Excel) 파일에 저장되었으며 자료 분석 시간을 최소화하고 정확도를 향상시키기 위해(윤효준, 박재현, 윤지운, 2019) 다음과 같은 전처리 단계를 거쳤다. 첫째, NLTK 자연어 처리 패키지를 활용하여 단어 단위로 텍스트를 분리하는 토큰화(tokenization) 과정을 거쳤다. 둘째, 분리된 단어들 중 명사, 형용사, 동사만 추출하는 표제어 추출 작업을 진행하였다. 셋째, ver, α, three, may 등과 같은 불용어(stop words)를 제거하는 정제 작업을 거쳤다.

3. 텍스트 마이닝 기반 토픽 모델링

텍스트 마이닝은 빅데이터 분석 방법 중 하나로 데이터 간 관계를 규명하는 과정을 통해 대량의 비구조화된 텍스트에서 새로운 지식을 추출하는 알고리즘이다(Li & Wu, 2010). 이러한 텍스트 마이닝은 크게 키워드분석과 토픽분석으로 나뉘며 토픽분석에는 의미연결망분석, 토픽모델링, 감성분석 등의 기법이 존재한다(서대호, 2019).

본 연구에서는 토픽 모델링 기법을 활용하였다. 토픽 모델링 기법은 논문의 초록을 분석해 주요 내용 추출 및 시간 흐름에 따른 주제 분포 변화를 파악하는데 적합하며 이러한 특성으로 인해 특정 학문분야 연구 동향을 탐색하는 연구에서 활용도가 높다(남춘호, 2016). 토픽 모델링의 모델 중에도 잠재의미분석(Latent Semantic Analysis, LSA) 알고리즘의 경우 문서 내 단어가 정규분포를 따라야 적용 가능하기 때문에 주제 추출에 용이하고 문서별 주제 비율, 단어가 주제에 포함될 확률 등을 파악(박자현, 송민, 2013)하는데 용이한 잠재디리클레할당(Latent Dirichlet Allocation, LDA) 알고리즘이 학계의 표준으로서 인정받고 있다(윤효준, 박재현, 윤지운, 전민수, 2021). 이에 본 연구에서는 Python 3.7 Package를 활용해 LDA 알고리즘 기반 토픽 모델링 기법을 적용하여 분석을 실시하였다.

토픽 모델링에서는 토픽의 수를 결정하기 위한

www.kci.go.kr

사전, 사후 지정 방식이 있는데 본 연구에서는 Gensim 라이브러리를 적용, 일관성 점수(Coherence Score)에 기준하여 통계적으로 가장 적합한 토픽 수를 사후 설정하였다(진주연 등, 2021). 그 결과 <그림 1>과 같이 일관성 점수가 가장 높은 지점을 최적의 토픽 수로 결정하여(안병은, 2020; 윤순욱, 김민철, 2020; 윤효준 등, 2021) 최종 6개의 토픽을 선정하였다.

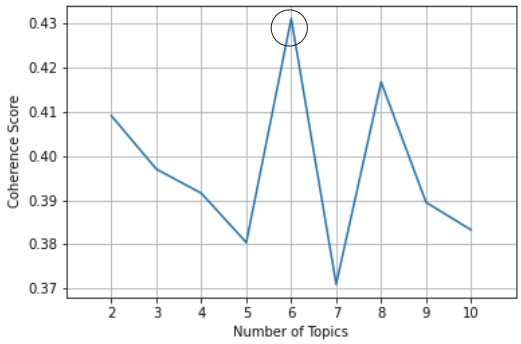


그림 1. Coherence Score에 따른 토픽 수 결정

Ⅲ. 연구결과 및 논의

1. 한국특수체육학회지의 단어 빈도 및 주제별 주요 키워드 분석

1) 한국특수체육학회지의 단어 빈도 분석 및 워드 클라우드 시각화

본격적인 토픽모델링 분석에 앞서 지난 29년간 어떠한 단어들이 연구의 주요한 키워드로 활용되었는지를 확인하기 위해 단어 출현 빈도 분석을 실시하였다. 한국특수체육학회지에 게재된 논문 영문초록 텍스트를 토큰화, 표제어 추출, 불용어 제거 등의 전처리 단계를 거쳐 빈도 분석한 결과 총 5843개의 단어가 확인되었으며 그 중 빈도수가 높은 상위 20개의 단어는 아래에 제시된 <표 1>과 같았다.

이와 같은 영문초록 내 출현 단어 빈도 분석 결과

표 1. 출현 빈도 상위 20개 단어

단어	n	단어	n
disability	2405	school	441
physical	2237	inclusive	439
sport	1341	body	403
exercise	1127	behavior	393
activity	1076	special	379
education	1073	self	353
program	1037	social	345
student	966	health	342
adpated	596	teacher	341
chilld	525	training	334

로 미루어 보았을 때 disability(2405회), physical(2237회), sport(1341), exercise(1127회), activity(1076회)등이 출현 빈도 상위권을 형성한 것은 특수체육학의 학문적 기반 및 특성이 투영된 결과로 볼 수 있다. 또한 education(1073회), program(1037회), student(966회), child(525회), school(441회), inclusive(439회) 등이 그 뒤를 이어 상위 빈출 단어로 등장한 것은 한국특수체육학회지 내에서 아동과 학령기 대상자가 연구 대상으로 많이 선정되었으며 교육 또는 특정 프로그램, 통합적 접근을 주제로 한 연구가 많았던 것으로 판단된다. 아동의 특성과 수준을 효율적으로 반영하여 체육 프로그램의 실제성과 효과성을 향상시키는 것(노형규, 2020)은 특수체육 현장의 핵심적인 가치다. 따라서 다양한 프로그램을 적용해 그 효과를 검증하고자 하였던 연구들이 지속적으로 이어졌던 것(홍양자, 1993; 장경호, 1998; 정이루리, 김기홍, 2003; 강유석, 맹혁주, 2007; 장창호, 2012; 임상훈, 김동민, 이용호, 2018; 김대경, 이현수, 2021)이 이러한 결과로 나타났다고 사료된다.

아래 <그림 2>는 한국특수체육학회지 영문초록 출현 빈도 상위 50개 단어를 워드 클라우드 방식으로 시각화한 결과다. 워드 클라우드(Word Cloud)는 문서의 키워드를 파악하는 데에 용이하며 높은 출현 빈도를 보이는 핵심적인 단어들을 시각화하여 데이터의 특징을 알 수 있게 해준다는 장점이 있다(김경진, 2018). 워드 클라우드에서는 출현 빈도에 따라 단어의 크기와 색 등이 상이하게 나타난

www.kci.go.kr



그림 2. 출현 빈도 상위 50개 단어의 워드 클라우드

다(유민수, 2021).

위와 같이 출현 빈도 기준 상위 50개 단어를 워드 클라우드 시각화하는 과정을 통해 출현 빈도 상위권에 위치하였던 단어들 중 출현 빈도 2000회 이상을 기록한 'disability', 'physical'과 같은 단어들은 가장 크고 굵은 글씨로 시각화되는 것을 확인할 수 있다. 또한 출현 빈도 1000회 내외를 기록한 'sport', 'exercise', 'activity', 'education', 'program', 'student' 등이 그 다음으로 크기와 굵기가 강조되어 표현되는 것을 볼 수 있다. 이러한 시각화 분석 과정을 통해 한국특수체육학회지 내에서 가장 많이 등장하는 상위 50개 단어를 이해하기 쉽게 집약적으로 파악하고 추출 데이터의 대략적인 특징을 확인할 수 있었다.

2) 토픽 모델링 분석을 통해 도출한 한국특수체육학회지의 주요 키워드 및 주제

LDA 알고리즘 기반 토픽 모델링 분석을 실시한 결과 한국특수체육학회지의 1993년 1권 1호에서부터 2021년 29권 4호까지 게재된 924개의 논문의 영문초록에 대한 총 6개의 토픽이 도출되었다. 토픽의 수는 위에서 밝힌 바와 같이 Coherence Score에 기준하여 선정하였다.

먼저 본 연구의 분석 대상으로 선정된 총 932개 논문에 대한 토픽별 개수와 비율을 살펴본 결과 'topic 4(227개, 24.5%)', 'topic 6(170개, 18.4%)', 'topic 2(168개, 18.2%)', 'topic 1(161개, 17.4%)', 'topic 5(106개, 11.6%)', 'topic 3(92개, 10%)'과 같은 순으로 나타나는 것을 확인할 수 있었다. 자세한 내용은 아래 제시된 <표 2>와 같다. 선정된 6개의 토픽에는 각 5개의 토픽별 주요 키워드가 추출되었으며 주요 키워드들의 특징 및 키워드 간 연관성을 고려하여 범주화된 주제로 토픽명을 지정하였다. 이에 대한 상세한 내용은 다음에 제시된 <표 3>을 통해 확인할 수 있다.

<표 2>와 <표 3>의 결과를 종합하여 미루어보았을 때 최기수와 우상연(2003), 박병도와 손천택(2007), 백재근과 전혜자(2011), 배규현, 김대연, 한민규(2013)의의 연구 등과 같이 '특수학교(급) 현장'에 관한 토픽이 가장 높은 비율을 차지하며 김

표 2. 토픽별 개수 및 비율

토픽	개수	비율
topic 1	161	17.4%
topic 2	168	18.2%
topic 3	92	10%
topic 4	227	24.5%
topic 5	106	11.6%
topic 6	170	18.4%
합계	924	100

표 3. 토픽모델링 분석 결과

연번	토픽	핵심어				
1	운동 프로그램 효과	exercise	physical fitness	program	balance	body fat
2	선수 사회적 지원	handicapped	support	social	athlete	facility
3	사회·심리학적 접근	physical activity	experience	female	self efficacy	gender
4	특수학교(급) 현장	physical education	teacher	inclusive	school	class
5	선수 트레이닝	player	training	technique	motion	speed
6	아동과 부모	child	behavior	parents	ability	student

www.kci.go.kr

권일과 김미량(2006), 임인선과 박현정(2007)의 연구 등과 같이 '아동과 부모'에 관련된 토픽 비율이 두 번째라는 것을 확인할 수 있다. '선수 사회적 지원'과 관련된 주제가 세 번째로 많은 비율을 차지하였으며 특정 프로그램을 적용하여 그 효과를 검증하는 연구들이 주를 이룬 '운동 프로그램 효과'에 관한 주제가 네 번째였다. 이외에도 '선수 트레이닝', '사회·심리학적 접근'에 관한 주제가 순서대로 그 뒤를 이었다. 이처럼 토픽 모델링을 활용한 분석의 경우 기존 연구 동향 분석에서 활용되어왔던 내용분석 방법과는 다르게 추출된 핵심어를 활용하여 전반적인 연구 주제와 경향을 한 눈에 쉽게 확인할 수 있다는 장점이 있다. 하지만 연구 방법, 연구 대상 등으로 구분하여 영역별 상세 내용과 세부적인 특징을 파악하기에는 다소 한계가 있다(장익영, 이원미, 2019).

2. 한국특수체육학회지 연구주제의 시계열분석

본 연구에서는 1993년부터 2021년까지 한국특수체육학회지에 게재된 논문을 바탕으로 구성한 6개의 토픽이 시간의 흐름에 따라 어떤 추이를 보이는지 파악하기 위해 연도별 논문의 발행 호 수와 발행 편수를 고려하여 시기를 구분하였다(장익영, 이원미, 2019). 이에 따라 시기를 총 6개 구간으로 구분하였으며 제1시기를 1993년~2002년, 제2시기를 2003년~2006년, 제3시기를 2007년~2010년, 제4시기를 2011년~2014년, 제5시기를 2015년~2018년, 제6시기를 2019년~2021년으로 지정하였다. 또한 시기별 연구 토픽의 추이를 확인하기 위해 최하단에 제시된 <표 4>와 같이 시기별 토픽 점유율을 산출하였다.

위 제시된 <그림 3>은 앞서 제시되었던 <표 4>의 점유율 시기별 토픽 점유율 데이터를 바탕으로 시기별 토픽 점유율의 변화 추이를 한 눈에 보기 쉽게 꺾은선 그래프로 시각화한 결과다. <표 4>와 <그림 3>의 내용을 종합하여 살펴보면 먼저 topic 1(운동 프로그램 효과)의 경우 전체 여섯 개의 토픽 중 네 번째로 높은 통합점유율(17.4%)을 보인다. 이러한 topic 1의 시기별 점유율을 살펴보면 제1시기에서부터 제3시기까지 꾸준히 20% 초중반대의 점유율을 유지하다가 제4시기에서 큰 폭으로 감소한 후 가장 근래인 제6시기 들어서는 7%의 점유율을 보이며 다시 한 번 더 큰 폭으로 감소하는 추이를 보였다. 이러한 경향은 한국특수체육학회지 내에서 효과 검증 연구가 주를 이루는 실험연구의 수가 점점 감소하는 동향을 보인다는 선행연구(이현수, 김대경, 윤석민, 2021)의 결과를 뒷받침하는 내용이라 볼 수 있다.

표 4. 시기별 토픽 점유율

토픽 \ 시기	제1시기 (1993~2002)	제2시기 (2003-2006)	제3시기 (2007-2010)	제4시기 (2011-2014)	제5시기 (2015-2018)	제6시기 (2019-2021)	전체
n	147	134	167	145	173	158	924
topic 1 운동 프로그램 효과	.224	.201	.246	.145	.162	.070	.174
topic 2 선수 사회적 지원	.211	.090	.132	.214	.214	.222	.182
topic 3 사회 · 심리학적 접근	.027	.045	.096	.138	.092	.190	.100
topic 4 특수학교(급) 현장	.204	.321	.198	.269	.249	.241	.245
topic 5 선수 트레이닝	.143	.127	.120	.069	.116	.120	.116
topic 6 아동과 부모	.190	.216	.210	.166	.168	.158	.184
합계	1	1	1	1	1	1	1

www.kci.go.kr

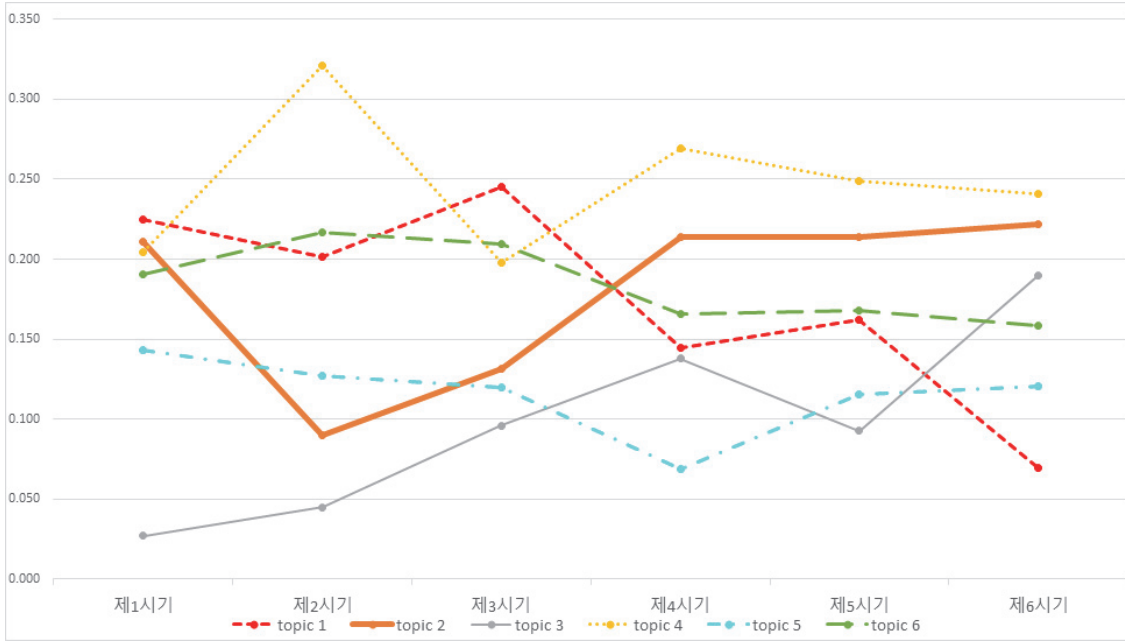


그림 3. 시기별 토픽 점유율 변화 추이

topic 2(선수 사회적 지원)의 경우 전체 토픽 중 세 번째로 높은 통합점유율(18.2%)을 나타냈으며 제1시기에서 제2시기로 가면서 큰 폭으로 감소했다가 제3, 제4시기로 들어서며 다시 크게 증가하였다. 제5, 6시기를 거치면서는 20% 초반대의 점유율을 꾸준히 유지하고 있으며 토픽간 시기별 점유율을 살펴보았을 때 제4,5,6시기에서 두 번째로 높은 점유율을 나타내고 있는 형세다. 이러한 경향은 기존의 장애인 및 체육 관련 법령에서 장애인체육 지원에 대한 명확한 한계가 존재하고(조재훈, 2021), 과거 일반체육의 구조를 답습해오던 장애인 체육 정책(윤석민, 2018)에 새로운 변화가 필요한 시기라는 부분에 있어 점점 더 많은 공감대가 형성되고 있음을 나타내는 결과라 판단된다.

topic 3(사회·심리학적 접근)은 토픽간 통합점유율 순위 여섯 번째(10%)인 것으로 나타났다. 시기별 점유율을 살펴보았을 때는 제1시기에서 제4시기까지 지속적으로 증가하다가 제5시기에서 처음으로 감소하고 제6시기 들어서는 다시 크게 증가하는 모습이다. 시기별 점유율에 있어 지속적으로 5~6위권에 머물다가 가장 최근인 제6시기 들어 3위로 올라섰는데 이러한 경향은 최근 사회심리학적 요인과 관련한 연구가 지속적으로 증가하고 있다는 선행연구(이현수 등, 2021) 결과와 동일한 결과라 볼 수 있다. 또한 운동 프로그램 효과 검증 연구와 같이 특정 주제에만 집중되어오던 특수체육 연구가 과거보다 더욱 다양한 주제군을 가지게 된 것(윤지운, 2017)으로 판단할 수 있으며 이는 topic 1(운동프로그램 효과)의 시기별 점유율이 근래에 들어서며 점차 감소하고 있는 것과도 관련 있는 부분이라 사료된다.

topic 4(특수학교(급)현장)은 전체 토픽 중 가장 높은 통합점유율(24.5%)을 보이고 있다. 시기별 점유율을 살펴보면 제1시기에서 제6시기까지 다른 토픽들에 비해 비교적 꾸준하게 상위권을 형성하고 있는 모습을 볼 수 있다. 특히 제2시기(2003-2006)에서는 약 32%에 달하는 점유율을 보이며 폭발적인 증가세를 보이고 있는 것을 확인할 수 있다. topic 4의 핵심어에 'physical education', 'teacher', 'inclusive', 'school', 'class'가 포함된 것으로 미루

www.kci.go.kr

어보았을 때 이는 2004년 교육인적자원부가 발표한 특수교육발전 종합계획 수정계획안의 영향일 것으로 판단된다. 수정계획안의 내용 중 장애학생과 비장애학생의 통합교육 환경 조성 및 방법 개선은 주요한 골자였다(이현수, 2005). 또한 특수교육대상자 수 자체의 확대뿐 아니라 통합교육을 희망하는 특수교육대상자가 증가하고 통합교육이 점차 확산(박인국, 이현수, 최수정, 2021)되어 자연스레 이에 관한 연구가 수반되고 있는 것이 시기별 점유율 순위에서 topic 4가 꾸준히 상위권을 형성하는 데에 영향을 미치는 요소일 것으로 사료된다.

topic 5(선수 트레이닝)은 전체 토픽 중 다섯 번째의 통합점유율(11.6%)을 나타내고 있다. 시기별 점유율은 제4시기를 제외하면 지속적으로 10% 초중반대의 점유율을 유지하며 꾸준히 출현하고 있다. 이러한 경향은 장애인 선수들에게 제공되는 체계적인 트레이닝과 우수 선수 발굴 및 양성 등에 있어서의 개선 필요성이 지속적으로 제기(김태홍, 박철형, 제갈윤석, 2015)되고는 있으나 일부 특정한 요인만을 고려하여 연구가 미비하게 진행(홍석만, 한민규, 2016)되고 있는 점을 드러낸다고 볼 수 있다.

topic 6(아동과 부모)는 통합점유율에 있어 두 번째(18.4%)로 높게 나타나고 있는 주제이며 시기별 점유율 순위에 있어서는 10% 중후반대에서 20%초반까지의 비율을 나타내어 꾸준히 중위권을 형성하고 있는 모습이다. topic 6에서는 장애아동의 문제행동에 관한 내용(박용천, 2003; 하창완, 2018) 또는 체육활동 참여에 따른 장애아동 부모들의 인식 및 만족도 등에 관한 연구(김의수, 오광진, 2000; 노형규, 오광진, 2003; 정세붕, 김지태, 2021) 등이 주를 이루었다. 이러한 연구 경향은 장애아동 교육에 있어 부모들의 역할과 욕구가 중요한 요소로서 인식(노형규, 오광진, 2003)되어왔으며 이를 반영하고자 하는 움직임이 지속적으로 나타나고 있는 것으로 보인다.

Ⅴ. 결론 및 제언

본 연구에서는 1993년부터 2021년까지 한국특수체육학회지에서 발행된 논문 중 924개 논문의 영문초록을 대상으로 LDA 알고리즘 기반 토픽 모델링 분석을 실시하여 연구주제의 특성을 파악하고 시간의 흐름에 따른 연구주제 동향 변화를 살펴보고자 하였다. 이에 대한 결론은 다음과 같다.

첫째, 토픽 모델링 분석 결과 한국특수체육학회지의 주요 토픽은 총 6개가 도출되었으며 점유율을 기준으로 하였을 때 '특수학교(급) 현장', '아동과 부모', '선수 사회적 지원', '운동 프로그램 효과', '선수 트레이닝', '사회·심리학적 접근' 순인 것으로 나타났다.

둘째, 지난 29년을 6개의 구간으로 구분하여 토픽별 점유율의 시계열적 변화를 살펴보았을 때 토픽별 변화 추이는 다음과 같았다. 먼저 topic 1(운동프로그램 효과)는 제1시기에서 제6시기까지 지속적으로 감소와 증가를 반복하는 경향을 보였으며 제3시기에서는 시기별 점유율 순위 1위(24.6%)를 차지하였다. 그러나 제4,5시기에서 중하위권으로 하락한 데 이어 제6시기에서는 점유율 순위 최하위(7%)를 기록하였다. topic 2(선수 사회적 지원)은 제1시기에서 제2시기로 가면서 대폭 감소했던 것을 제외하면 꾸준히 증가하는 추이를 보였다. 시기별 점유율로 보면 토픽들 중 두 번째로 높은 점유율을 기록했던 제6시기(22.2%)에서 제일 높은 점유율을 기록하였다. topic 3(사회·심리학적 접근)의 경우 제1시기부터 제4시기까지 지속적으로 증가하다가 제5시기에서 감소하였고 제6시기에서 다시 큰 폭으로 증가하는 추이를 보였다. 시기별 점유율 순위에 있어서는 지속적으로 하위권을 형성하다 제6시기 들어 19%의 점유율을 보이며 중상위권으로 반등하는 모습이었다. topic 4(특수학교(급) 현장)은 제1시기에서 제4시기까지 증가와 감소를 반복하였으며 제4시기부터 제6시기까지는 25% 내외의 점유율을 유지하였

www.kci.go.kr

다. topic4의 시기별 최고 점유율은 제2시기에서 기록한 32.1%였다. topic 5(선수 트레이닝)은 6.9%로 떨어졌던 제4시기를 제외하면 큰 폭의 변화 없이 꾸준히 10% 초중반대의 점유율을 유지하였다. 시기별 최고 점유율은 14.3%를 차지했던 제1시기로 나타났다. topic 6(아동과 부모)는 제1시기부터 제3시기까지 20% 내외의 점유율을 보였으나 제4시기부터 감소하여 15~16%대의 점유율을 나타내었다. 또 토픽간 시기별 점유율 순위 2위를 차지했던 제2시기에서 최고 점유율(21.6%)을 보였다.

한편 토픽별 전체 점유율 순서와는 달리 가장 근래의 시기인 제6시기(2019년~2021년)을 기준으로 하였을 때는 topic 4(특수학교(급) 현장)이 가장 높은 점유율을 보였고 topic 2(선수 사회적 지원)과 topic 3(사회·심리학적 접근)이 각각 두 번째, 세 번째로 높은 점유율을 보였다. 또 그 뒤를 topic 6(아동과 부모), topic 5(선수 트레이닝), topic 1(운동 프로그램 효과)가 이어나가는 모습이었다. 이러한 결과는 특수체육 내에서 근래 들어 어떠한 부분이 조명 받고 있으며 많은 공감을 얻고 있는지 사회적 흐름을 파악 및 예측해볼 수 있는 결과라고 볼 수 있다. 또한 제1시기와 제6시기의 토픽별 점유율을 비교했을 때 topic 1(운동 프로그램 효과)의 점유율이 22.4%에서 7%로 1/3가량 줄어들고 topic 3(사회·심리학적 접근)의 점유율이 2.7%에서 19%로 대폭 증가한 것은 눈여겨볼만한 결과다. 이는 특수체육에 대한 논의가 단순히 운동 프로그램의 효과 검증과 같이 재활, 의료적 관점에서의 접근에서 이제는 더 나아가 사회적 모델 관점으로 발전하여 특수체육 연구 분야의 폭과 범위가 더욱 확장(최윤소, 이예슬, 2021)되고 있다는 반증으로 생각된다.

위 내용을 통해 확인할 수 있었듯 토픽 모델링 분석 기법은 핵심어에 기반하여 한 분야 내에서 전반적으로 어떠한 연구 주제들이 다루어지고 있는지 그 경향을 파악하고 점유율 등의 데이터를 분석 및 시각화하기에 용이하다는 장점이 있다. 또한 해당 자료를 바탕으로 학문적 논제를 제시하고 다학제적 연구 접근 가능성을 확인하는 등의 의미를 가진다. 하지만 연구방법, 연구대상 등과 같이 더욱 세부적으로 분류하며 특성을 파악하는데 있어서는 제한이 있다. 또한 토픽별 논문 군집과정에서 영문초록 내 데이터 출현 빈도에 기반한 단순 기계적 방법으로 각 논문들의 성격을 판단하였기에 각 논문별 연구자들의 의도를 적확히 반영하는 것에는 어려움이 있다. 더불어 본 연구에서는 웹 크롤링을 활용하는 분석 방법의 특성상 영문초록이 존재하지 않는 27개 논문을 분석대상에서 제외하였고 국문초록 또한 분석 대상으로 설정하지 않았다는 한계가 있었다. 이러한 점을 고려하여 후속연구에서는 자료수집 범위를 논문의 전문으로 확대하고 내용분석 등의 방법을 함께 활용하여 분석 영역을 연구 방법과 대상으로 확장하는 등 토픽 모델링의 방법론적 한계를 보완하는 것이 필요하다 판단된다.

참고문헌

강유석, 맹혁주(2007). 농구활동 프로그램이 정신지체학생의 예측타이밍에 미치는 영향. **한국특수체육학회지, 15**(1), 1-20.

김기룡, 김삼섭(2016). 텍스트마이닝을 이용한 국내 통합교육 연구 동향 분석. **지체·중복·건강장애연구, 59**(4), 133-157.

김경진(2018). 텍스트마이닝 기법을 활용한 하계패럴림픽 해외 연구 동향 분석. **한국특수체육학회지, 26**(2), 33-45.

김권일, 김미량(2006). 정신지체아동의 신체활동 참여기간에 따른 부모의 양육스트레스, 생활만족의 구조방정식 모형 비교. **한국특수체육학회지, 14**(4), 197-213.

김동남, 홍금석, 고진환, 전정환(2017). 한국의 항공우주 분야 연구동향 분석. **산업혁신연구, 33**(4), 193-226.

김대경, 이현수(2021). 지체장애노인의 건강생성 신체활

www.kci.go.kr

동 프로그램이 통합성 및 긍정심리자본에 미치는 영향. **한국특수체육학회지, 29**(2), 233-243.

김지은, 백순근(2016). 텍스트 빅데이터 분석 기법을 활용한 대학구조개혁 평가의 쟁점 분석. **아시아교육연구, 17**(3), 409-436.

김태홍, 박철형, 제갈윤석(2015). 맞춤형 트레이닝이 장애인 엘리트선수들의 체력수준에 미치는 영향. **한국웰니스학회지, 10**(4), 355-367.

남춘호(2016). 일기자료 연구에서 토픽모델링 기법의 활용가능성 검토. **비교문화연구, 22**(1), 89-135.

노형규(2020). **발달장애 아동 체육 프로그램의 이해와 실제**. 서울: 레인보우북스.

노형규, 오광진(2003). 장애아동의 신체활동 프로그램 참여에 대한 부모의 인식. **한국특수체육학회지, 11**(1), 1-14.

문주영(2009). 텍스트마이닝을 이용한 비서학의 지적구조 이해. **비서학논총, 18**(1), 83-98.

박병도, 손천택(2007). 지체부자유학교 교사의 체육수업 어려움 해결하기. **한국특수체육학회지, 15**(3), 1-26.

박용천(2003). 단체줄넘기운동이 정서장애아의 문제행동에 미치는 영향. **한국특수체육학회지, 11**(2), 101-117.

박인국, 이현수, 최수정(2021). 장애학생과의 스포츠클럽 수업이 비장애학생의 상호협력행동, 윤리적민감성, 공동체 의식에 미치는 영향. **한국특수체육학회지, 29**(2), 21-34.

박자현, 송민(2013). 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. **정보관리학회지, 30**(1), 7-32.

반현정, 강병남, 김학선(2021). 텍스트 마이닝 기반 토픽 모델링을 활용한 한식 조리 연구 동향 분석. **산업혁신연구, 37**(3), 141-163.

배규현, 김대연, 한민규(2013). 지적장애학생의 학교생활 내 휴식시간과 점심시간 신체활동과 비신체활동 내용분석. **한국특수체육학회지, 21**(1), 29-40.

백재근, 전혜자(2011). 통합체육수업과 특수학급 체육수업 시간 내 지적장애학생의 신체활동 수준 비교분석. **한국특수체육학회지, 19**(1), 103-114.

서대호(2019). **잡아라! 텍스트 마이닝 with 파이썬**. 서울: 비제이퍼블릭.

안병은(2020). **비지도 단어 임베딩 기반 토픽 모델링 상세 주제 변화 탐지**. 미간행 석사학위논문. 서울대학교 공학전문대학원.

유민수(2021). **최근 20년 체육학 학술지에 게재된 청소년연구 토픽분석**. 미간행 석사학위논문. 한국체육대학교 사회체육대학원.

윤석민(2018). 장애인생활스포츠 활성화를 위한 추진과제: 한국과 미국의 장애인스포츠 관련법 비교. **스포츠사이언스, 36**(1), 77-85.

윤순욱, 김민철(2020). LDA 기법을 이용한 미세먼지 이슈의 토픽모델링 분석. **에너지공학, 29**(2), 23-29.

윤지운(2017). 지식네트워크 분석을 활용한 특수체육의 연구동향 분석. **한국특수체육학회지, 25**(1), 121-133.

윤홍근(2013). 문화산업에서 빅데이터의 활용방안에 관한 연구. **글로벌문화콘텐츠, (10)**, 157-179.

윤효준, 박재현, 윤지운(2019). 비정형 텍스트 자료에서 잠재 정보 추출을 위한 토픽모델링 소개: 치매관련 신체활동 뉴스 기사의 이슈 분석. **체육과학연구, 30**(3), 501-512.

윤효준, 박재현, 윤지운, 전민수(2021). 토픽모델링을 적용한 코로나19 확산에 따른 스포츠영역 이슈 분석. **한국체육측정평가학회지, 23**(1), 11-20.

이경옥, 하지연(2000). 특수체육의 연구동향에 관한 분석. **한국특수체육학회지, 8**(2), 17-34.

이현수(2005). 지체장애학생과의 통합체육수업이 비장애학생의 장애학생에 대한 인식에 미치는 영향. **한국스포츠교육학회지, 12**(1), 207-221.

이현수, 김대경, 윤석민(2021). 한국과 미국의 특수체육 연구동향 비교분석(2016-2020): APAE와 APAQ를 중심으로. **한국특수체육학회지, 29**(4), 43-52.

이현수, 박수경, 박진우, 윤석민, 최승오(2011). 한국과 미국의 특수체육 연구동향 비교분석(2006-2010): APAE와 APAQ를 중심으로. **한국특수체육학회지, 19**(3), 63-74.

임상훈, 김동민, 이용호(2018). 변형된 골프 프로그램(Adapted golf program)이 지적장애 학

www.kci.go.kr

생의 시지각 및 주의집중에 미치는 영향. **한국특수체육학회지, 26**(1), 1-16.

임인선, 박현정(2007). 장애아동을 무용요법에 참여시키는 어머니들의 참여동기와 지속경험 이해. **한국특수체육학회지, 15**(1), 45-63.

장경호(1998). 학습장애학생의 학교생활태도와 이미지 변화를 위한 스케이트보드 프로그램의 효과. **한국특수체육학회지, 6**(1), 1-16.

장익영, 이원미(2019). 토픽모델링을 적용한 한국스포츠사회학회지의 연구주제 분석. **한국스포츠사회학회지, 32**(2), 20-35.

장창호(2012). 감각운동통합 프로그램이 뇌성마비 아동의 운동제어에 미치는 영향. **한국특수체육학회지, 20**(2), 107-117.

전채남, 서일원(2013). 빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로. **마케팅논집, 21**(2), 181-203.

정경환, 원영신, 구송광(2015). 특수체육 연구동향 분석: 2011-2014. **한국특수체육학회지, 23**(1), 79-90.

정복자, 홍양자(2006). 특수체육학의 최근 연구동향에 관한 분석. **한국여성체육학회지, 20**(1), 65-72.

정세붕, 김지태(2021). 특수체육기관 선택속성과 학부모 만족도 및 지속행동의도의 인과관계: 장애 자녀를 둔 학부모를 대상으로. **한국특수체육학회지, 29**(1), 27-40.

정이루리, 김기홍(2003). 감각계 훈련 프로그램이 다운증후군 아동의 평형성에 미치는 영향. **한국특수체육학회지, 11**(3), 13-26.

조재훈(2021). 장애인체육 지원을 위한 법률 제정의 필요와 방향. **한국특수체육학회지, 29**(4), 1-17.

진주연, 윤지은, 김은혜, 김지영(2021). 토픽모델링 및 의미연결망을 활용한 지적장애인 대상 특수체육 연구의 동향 분석. **한국특수체육학회지, 29**(1), 103-121.

최기수, 우상연(2003). 일반학급 학생과 특수학급 학생의 체력요인 비교. **한국특수체육학회지, 11**(1), 93-108.

최승권(2018). 특수체육 지도자를 위한 양성 과정과 진로. **한국특수체육학회지, 26**(4), 1-11.

최윤소, 이예슬(2021). 체육학에서의 장애인 체육 연구 동향 분석. **한국융합과학회지, 10**(3), 35-51.

하창완(2018). 뉴스포츠 프로그램 중심의 통합 체육 교실 운영이 정서·행동장애 청소년의 문제행동 및 사회적 기술에 미치는 영향. **한국특수체육학회지, 26**(1), 17-31.

홍석만, 한민규(2016). 휠체어육상선수들을 위한 체력 트레이닝 프로그램 모형 개발. **한국특수체육학회지, 24**(3), 13-27.

홍양자(1993). 정서장애 아동을 위한 체육 프로그램 개발에 관한 연구. **한국특수체육학회지, 1**(1), 49-60.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*, 993-1022.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision support systems, 48*(2), 354-368.

Sheldon, P. J. (1991). An authorship analysis of tourism research. *Annals of Tourism Research, 18*(3), 473-484.

논문제출일 : 2022년 4월 26일
게재승인일 : 2022년 6월 22일

www.kci.go.kr

Exploring Research Trends of Korean Journal of Adapted Physical Activity using Text Mining-based Topic Modeling

Gun-hwan Bae*(Korea National Sport University / Doctoral student)*·
Hyoung-kuy Roh**(Korea National Sport University / Professor)*

Abstract

The purpose of this study was to grasp the overall research trends of the Korean Journal of Adapted Physical Activity by identifying the themes of the studies published in the Journal of the Korean Journal of Adapted Physical Activity for the past 29 years, and analyzing the issues, trends, and characteristics of each period. To this end, text mining-based topic modeling analysis was performed on 924 papers with English abstracts among 951 papers published in the Korean Journal of Adapted Physical Activity from 1993 to 2021. LDA(Latent Dirichlet Allocation) algorithm-based topic modeling was applied using Python 3.7 Package. As a result of the analysis, a total of 6 topics were derived: 'exercise program effect', 'athlete social support', 'social/psychological approach', 'special schools(class) field', 'athlete training', 'children and parents'. In addition, in order to understand the research trends according to the time series, the period was divided into 6 sections and analyzed. Looking at the trend of change in topic share by period, the topic of 'exercise program effect' maintained a constant share (22.4%, 20.1%, 24.6%) from the 1st period to the 3rd period, and after a significant decrease in the 4th period (14.5%) The 5th period (16.2%) showed a slight increase, and the most recent period, the 6th period (7%), showed a significant decrease. The topic of 'athlete social support' decreased significantly from the 1st period (21.1%) to the 2nd period (9%), then increased again through the 3rd and 4th periods (13.2%, 21.4%), and then increased again in the 5th period (21.4%)) increased slightly in the 6th period (22.2%) after the stagnation period. The topic of 'social/psychological approach' showed an inclination of steadily increasing from the 1st period (2.7%) to the 4th period (13.8%), decreasing in the 5th period (9.2%), and then rebounding sharply again in the 6th period (19%). The topic of 'special schools(class) field' showed characteristics such as tendency to form a high-ranking share by period, repeating an increase or decrease in general, and in particular, an explosive increase in the 2nd period (32.1%), and ranked 1st in the overall period's combined share (24.5%). Except for the 4th period (6.9%), ‘athlete training’ continued to maintain a share in the early to mid-range of 10%, and showed a tendency to appear steadily. Lastly, the 'children and parents' topic maintained a share of around 20% from the 1st to the 3rd period (19%, 21.6%, 21%), and showed a share of around 16% from the 4th to the 6th period (16.6%, 16.8%, 15.8%), steadily forming the middle rank. These results are meaningful in that we can identify and predict the future direction of adapted physical activity by analyzing and arranging the current status and trends of research in the Korean Journal of Adapted Physical Activity. It is believed that the results of this study can be used as basic data for future research for the expansion and development of adapted physical activity research areas.

Key word : Korean Journal of Adapted Physical Activity, research topic, research trend analysis, text mining, topic modeling

* rhk0019@knsu.ac.kr

www.kci.go.kr