



ISSN: 2733-7146 © 2022 KODISA & JRPE
 JRPE website: <https://acomis.kisti.re.kr/JRPE>
 doi: <http://dx.doi.org/10.15722/jrpe.3.1.202203.7>

Overseas Research Trends Related to 'Research Ethics' Using LDA Topic Modeling*

Woo-Ryeong YANG¹, Hoe-Chang YANG²

Received: January 15, 2021. Revised: November 29, 2021. Accepted: December 05, 2021.

Abstract

Purpose: The purpose of this study is to derive clues about the development direction of research ethics and areas of interest which has recently become a social issue in Korea by confirming overseas research trends. **Research design, data and methodology:** We collected 2,760 articles in scienceON, which including 'research ethics' in their paper. For analysis, frequency analysis, word clouding, keyword association analysis, and LDA topic modeling were used. **Results:** It was confirmed that many of the papers were published in medical, bio, pharmaceutical, and nursing journals and its interest has been continuously increasing. From word frequency analysis, many words of medical fields such as health, clinical, and patient was confirmed. From topic modeling, 7 topics were extracted such as ethical policy development and human clinical ethics. **Conclusions:** We founded that overseas research trends on research ethics are related to basic aspects than Korea. This means that a fundamental approach to ethics and the application of strict standards can become the basis for cultivating an overall ethical awareness. Therefore, academic discussions on the application of strict standards for publishing ethics and conducting researches in various fields where community awareness and social consensus are necessary for overall ethical awareness.

Keywords : Ethics, Research Ethics, Overseas research trends, LDA, Topic Modeling.

JEL Classification Code: C10, C49, C60, H80.

1. Introduction

윤리에 대한 기준은 선진국 대열에 서게 된 우리나라에서도 과거와는 다르게 개인 차원뿐 아니라 집단 및 조직 차원까지 확대되었으며, 인간 생활과 직·간접적으로 관련이 있는 다양한 분야에서 엄격하게 적용되고 있는 것이 현실이다. 예를 들어 속칭 '라면

상무 사건'이나 '땅콩 회항 사건', 대기업 경영진의 '운전기사 폭행·폭언사건', '성추행 사건', '치즈 강매 및 보복 출점' 등은 개인의 몰락뿐 아니라 소비자의 불매운동 등으로 인해 브랜드 가치가 훼손되고 소속 기업의 주가가 폭락하는 등 과거와는 다르게 국민으로부터 엄정한 단죄의 철퇴를 맞게 되었다. 또한 유명인들의 학폭사태, 군부대 상관의 다양한 갑질,

* This paper was presented at ICBE2022 and revealed that it was reviewed by three anonymous reviewers.

**This study was supported by the research grant of the KODISA Scholarship Foundation in 2021.

¹ First Author. Integrated Course of Master & Doctoral, Dept. of Business Informatics, Hanyang University, Korea. Email: wooryeong325@gmail.com

² Corresponding Author. Assistant Professor, Department of Distribution Management, Jangan University, South Korea. Email: pricezzang@jangan.ac.kr

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

정치인들의 개인적인 부정부패뿐 아니라 공사집단이나 기업까지 비윤리적이라는 지적에 누구도 자유로울 수 없는 분위기로 확대되었다. 본 연구에서 관심을 갖고 있는 연구윤리 또한 최근까지 사회적으로 큰 파문을 일으키고 있는 분야이다. 과거 정치인들이 어려움에 처하는 일반적인 이슈가 '성실납세 여부', '군 의무복무 준수 여부', '일반 법규 준수 여부'가 주를 이루었다면 최근에 등장하는 기준으로 '학위논문 표절 여부'가 당연히 등장하게 된 것도 연구윤리에 대한 많은 이들의 관심뿐 아니라 공인의 윤리적 기준에 대한 도덕률로 작용한 탓으로 보인다. 그러나 지금까지 우리나라 연구윤리에 대한 학계의 연구결과는 미미한 것으로 확인되었다. 2021년 11월 18일까지 scienceON(www.scienceon.kisti.re.kr)에 등록된 연구윤리 관련 국내논문은 2000년부터 2021년까지 총 106편인 것으로 나타났다. 이 중에서 '한국생명윤리학회'의 '생명윤리(Journal of the Korea Bioethics Association)'가 19편(17.9%)으로 가장 많았으며, '한국유통과학회'의 '연구윤리(Journal of Research and Publication Ethics)'가 10편(9.4%), '유통과학연구(Journal of Distribution Science)'가 2편(1.9%) 정도인 것으로 나타났다.

본 연구는 최근 사회적으로 이슈가 되고 있는 연구윤리에 대한 해외 연구동향을 확인함으로써 우리나라 연구윤리에 대한 발전방향과 관심영역에 대한 단서들을 도출하여 건전하고 윤리적인 연구문화를 정착시키는데 기여하고자 하는 것이 주요 목적이다. 또한 해외 연구동향을 확인을 통해 출판윤리뿐 아니라 연구윤리에 대한 방향성과 사회적 이견조율에 대한 단서들을 도출하는 것도 하나의 목적이다. 이를 위해 본 연구에서는 한국과학기술정보연구원의 scienceON에 탑재된 해외 논문 중 키워드에 'research ethics'를 포함한 논문 총 3,494편 중 유효한 논문 2,661편을 추출하여 분석에 활용하였다. 분석은 R 4.1.1 패키지를 활용하여 키워드 빈도분석, 워드클라우드,

단어 연관분석 및 LDA 알고리즘을 활용한 토픽모델링 분석을 실시하고 결과를 시각화하였다.

본 연구의 결과는 우리나라 교육기관과 연구기관이 연구윤리에 대한 기준을 수립하는데 다양한 단서들을 제공해줄 것으로 기대한다. 한편 해외 연구동향을 확인함으로써 연구윤리의 방향과 연구윤리의 대상에 대해서도 확인할 수 있는 기회를 제공해줄 것으로 기대하였다.

2. Literature Review

2.1. Research Ethics

윤리는 우리 자신과 다른 사람의 행동에 대한 기대를 지배하는 일련의 규칙이다(Mackinnon, 2009). 따라서 연구윤리는 과학적 연구를 수행하고 발표하는 방법에 대해 안내하는 일련의 윤리적 지침으로 다면적인 윤리성, 합법성과 공동체의 구성원으로서 행정적 문제를 포함한 연구자의 행동 기준을 규율할 뿐 아니라 책임있는 연구를 수행하기 위한 지침이다(Adhikari, 2020).

우리나라의 '연구윤리'와 관련된 기준은 앞에서 제시한 '연구윤리 확보를 위한 지침' 및 '국가연구개발혁신법'과 같은 법령과, 국가생명윤리정책원(KoNIBP)의 '연구윤리규정'뿐 아니라 한국환경연구원의 '연구윤리규정' 등 각 기관별로 규정화하고 있으며, 대학들도 각자의 연구윤리와 관련된 규정을 제정하여 운영하고 있다.

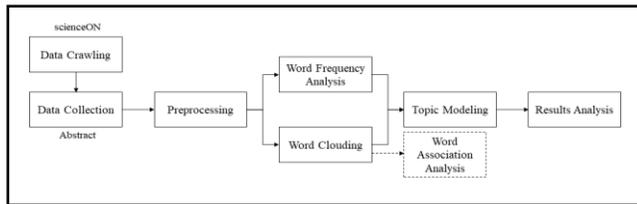
2.2. LDA(Latent Dirichlet Allocation) Topic Modeling

텍스트마이닝 기법 중 하나인 토픽모델링은 비구조화된 텍스트 자료들의 집합에서 의미있는 주제(토픽)를 추출해주는 확률적 모델 알고리즘(Blei, Ng, & Jordan, 2003)으로, 유사한 의미를 가지는 키워드들을 연결하는 방식으로 토픽을 분석하거나 추론하는 기법이다(Yang, in

press). 토픽모델링은 방대한 양의 문서집합에서 주요 토픽을 추출하고 각 토픽에 대응되는 문서를 식별하여 제공하며, 하나의 문서가 하나의 토픽으로만 할당되는 일반적인 군집화(dustering) 기법과는 다르게 하나의 문서가 여러 토픽에 동시에 대응될 수 있기 때문에 현실 세계의 모델링에 보다 적합한 기법으로 평가받고 있다(Kim, Lee, Choi, & Wong, 2017) .

2.3. Research Procedure

본 연구의 수행을 위해 한국과학기술정보연구원의 scienceON 에 탑재된 해외 논문을 대상으로 웹 크롤링을 실시하였다. 구체적인 연구방법과 연구절차는 아래 <그림 1>에 제시하였다.



Source: own

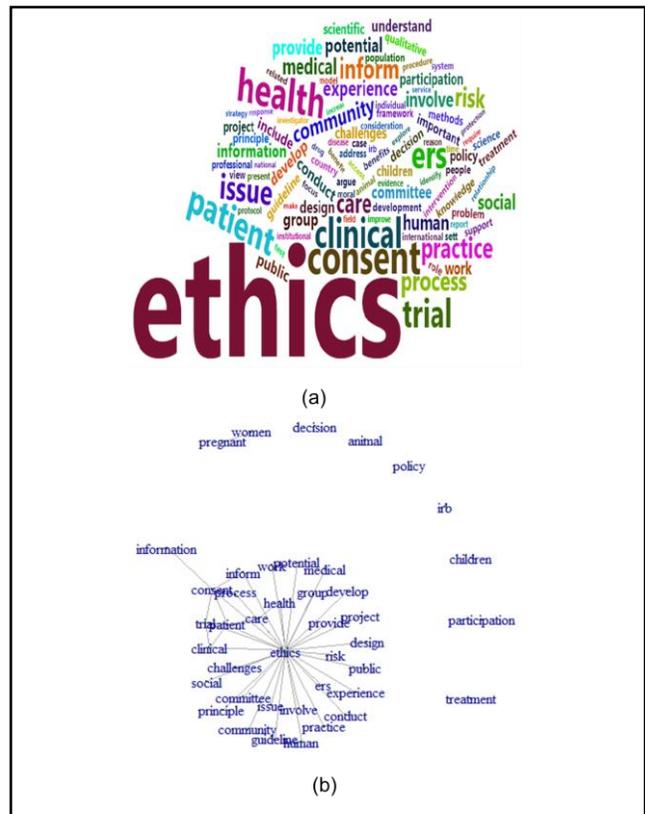
Figure 1: Research Procedure and Research Methods

3. Research Results

3.1. Word Frequency Analysis and Word Clouding

키워드 빈도분석 결과 윤리(ethics)를 제외하고는 동의(consent), 건강(health), 임상(clinical), 환자(patient), 유럽호흡기학회(ERS: European Respiratory Society), 의료(medical) 등 의학 분야의 단어들이 다수 출현하고 있는 것을 알 수 있었다. 한편, findAssocs함수를 이용하여 '윤리(ethics)'와 단어간 상관관계수가 0.3 이상인 단어를 분석한 결과 위원회(committee)만이 0.34로 나타났으며, 0.2 이상인 단어와의 상관관계를 확인한 결과 이슈(issue)가 0.29, 가이드라인(guideline)이 0.20으로 나타나 '윤리-이슈'와 '윤리-가이드라인'의 관계가 있는 것으로 확인되

었다. 단어 간의 연관성을 확인하기 위해 tdm.edgelist를 이용하여 가독성을 위해 연결빈도 1,500회를 적용하여 단어연관분석을 실시한 결과 윤리(ethics)에 대해 위험(risk), 건강(health), 돌봄(care) 등의 단어들이 방사선형으로 연관이 있는 것으로 나타났다. 이외에 여성들(women)과 임신(pregnant)의 연관이 있으며, 임상시험심사위원회(IRB: Institutional Review Board), 정책(policy), 동물(animal), 치료(treatment) 등이 별도로 연구되는 것으로 확인되었다. 키워드분석 결과를 워드클라우드링(word clouding)으로 시각화한 결과(a)와 단어 연관 분석을 시각화한 결과(b)는 <그림 2>에 제시하였다.



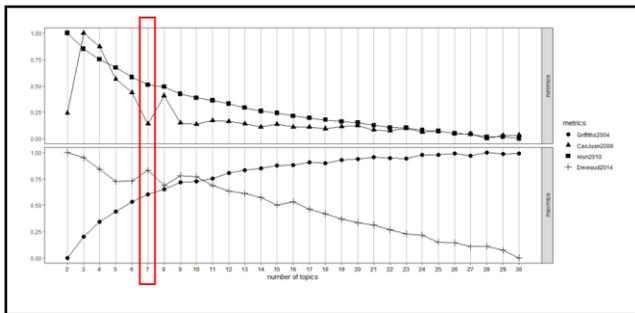
Source: own

Figure 2: Research Procedure and Research Methods

3.2. Topic Modeling

토픽 수(k)는 연구자들이 토픽모델링을 통해 결과 도출에 가장 적절하다고 판단한 숫자로 결정하는

것(Nahm, 2016)으로 특정한 기준이 없는 것이 사실이다. 그러나 토픽 수 결정에 연구자의 주관이 개입되는 것을 최소화하기 위해 다양한 방법이 활용되고 있다(Yang, in press). 본 연구에서는 Griffiths & Steyvers(2004), Cao, Xiz, Li, Zhang, & Tang(2009), Arun, Suresh, Veni Madhavan., & Narasimha Myrthy(2010), Deveaud, SanJuan, & Bellot(2014)가 제안한 방법을 R패키지의 FindTopicsNumber 명령어를 활용하여 토픽 수를 도출(Yang, in press)한 결과 토픽 수는 7개가 적합한 것으로 판단하였다.

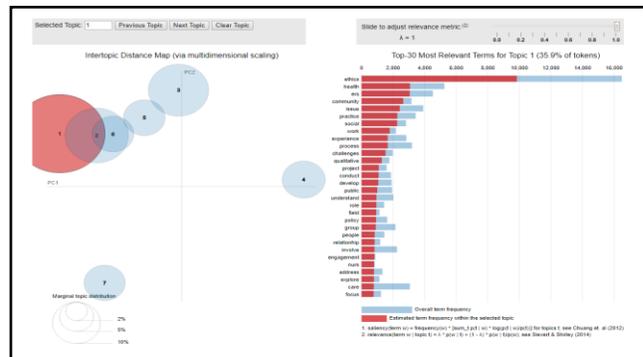


Source: own
Figure 3: Analysis results of determine the number of topics

LDA 토픽모델링 결과 각 토픽의 상위 15개 단어를 통해 각 주제별 토픽 명을 결정한 결과는 <표 1>에 제시하였으며 이를 LDAvis로 시각화한 결과는 <그림 4>에 제시하였다. LDAvis 결과는 토픽 1과 토픽 2, 토픽 2와 토픽 6이 상당히 중첩되어 있으며 토픽 1, 2, 3, 5, 6이 가까이 분포된 것을 알 수 있다. 이런 결과는 윤리정책 개발과 인간 임상윤리, 인간 임상윤리와 책임 출판이 상당히 연관성있게 연구되고 있음을 알 수 있으며, 전반적인 연구가 윤리정책 개발, 인간 임상윤리, 고지에 의한 동의, 전자 & 생명윤리, 책임 출판에 집중(84.5%)되어 있음을 확인할 수 있다. 한편, 토픽 4의 임신과 질병 처치, 토픽 7의 동물실험 윤리는 별도의 영역으로 연구되는 것을 알 수 있었다.

Table 1: Results of Topic Modeling

# of IDM	Ratio	Topic name	Top 15 words in each topic
1	35.9	ethical policy development	ethics, health, ers, community, issue, practice, social, work, experience, process, challenges, qualitative, project, conduct, develop
2	17.6	human clinical ethics	ethics, trial, clinical, consent, risk, inform, issue, health, patient, human, care, principle, potential, medical, ers
3	15.9	Informed consent	patient, consent, care, children, participation, experience, decision, parents, inform, clinical, group, understand, process, family, medical
4	8.1	pregnancy and disease treatment	trial, women, hiv, pregnant, patient, risk, clinical, disease, treatment, test, vaccine, rate, drug, intervention, population
5	7.6	Genes & Bioethics	health, consent, information, genetic, biobank, shar, inform, privacy, potential, ers, community, test, public, provide, material
6	7.5	Responsible publication	ethics, committee, publication, irb, medical, approval, ers, report, institution, country, c, protocol, include, train quality
7	7.4	animal experimentation ethics	animal, human, moral, ethics, scientific, science, embryo, scientist, technology, medical, bioethics, experiment, public, cell, stem



Source: own
Figure 4: Results of LDAvis

4. Results and Discussion

본 연구는 연구윤리에 대한 해외 연구동향을 확인함으로써 우리나라 연구윤리에 대한 발전방향과

관심영역에 대한 단서들을 도출하고, 출판윤리뿐 아니라 연구윤리에 대한 방향성과 사회적 의견 조율에 대한 단서들을 도출함으로써 건전하고 윤리적인 연구문화를 정착시키는데 기여하고자 하는 것이 주요 목적이다. 이를 위해 scienceON에 탑재된 총 2,661편의 논문초록을 분석한 결과와 시사점은 다음과 같다. 첫째, 분석한 논문들은 의학, 바이오, 제약, 간호 등의 학술지에 다수 게재된 것으로 확인되었으며, 시간이 지날수록 연구윤리에 대한 학술지와 연구자들과 관심이 높아지고 있는 것으로 확인되었다. 둘째, 단어 빈도분석결과도 동의, 건강, 임상, 환자 등 의학관련 분야의 단어들이 다수 출현함을 알 수 있었으며, 단어 연관분석 결과는 윤리와 관련하여 다양한 이슈들이 함께 연구되고, 여성과 임신, 정책, 동물 등이 별도로 연구되고 있는 것으로 나타나 우리나라에서 이슈가 되고 있는 출판 관련 윤리보다 다양한 분야의 윤리적 이슈를 연구하고 있는 것을 확인할 수 있었다. 마지막으로 토픽모델링 결과 윤리정책 개발, 인간 임상윤리 등 7개의 토픽을 추출하였다.

본 연구 결과를 통해 우리나라 관련 기관들이 연구윤리의 발전을 위해 정부 및 관련 기관이 연구윤리에 대한 가이드라인을 제시함에 있어 보다 근본적인 윤리적 측면을 고려하여야 함을 제안한다. 또한, 연구기관 및 학회의 경우 보고서나 학술지 발간을 통해 획득할 수 있는 비윤리적 사례를 공개하고 출판윤리와 관련된 다양한 영역에 대한 정보를 공유할 수 있는 노력을 아끼지 말아야 할 것이다. 마지막으로, 연구자는 연구수행 전 해당 분야의 연구윤리원칙과 세부 내용을 충분히 점검한 후 연구를 수행해야 할 것이다

References

- Adhikari, S. (2020). Research Ethics: Definition, Principles and Advantages. In AI ROI: Computer Vision whitepaper 2021. (<https://www.publichealthnotes.com/research-ethics-definition-principles-and-advantages/>)
- Arun, R., Suresh, V., Veni Madhavan. C. E., & Narasimha Myrthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In: Zaki M.J., Yu J.X., Ravindran B., Pudi V. (eds) *Advances in knowledge discovery and data mining*. PAKDD 2010. *Lecture notes in computer science*, vol 6118. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13657-3_43
- 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.
- 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.
- Kim, N. G., Lee, D. H. Choi, H. C., & Wong, W. X. S. (2017). Investigations on techniques and applications of text analytics. *The Journal of Korean Institute of Communications and Information Sciences(JKICS)*, 42(2), 471-492.
- Mackinnon, B. (2009). *ETHICS: Theory and contemporary issues*(6th Ed.). CA: Wardworth
- Yang, H. C. (in press). Topic Modeling Analysis of Franchise Research Trends Using LDA Algorithm. *The Korean Journal of Franchise Management.*, 12(4).