

텍스트 마이닝을 활용한 재생에너지 연구 동향: SCOPUS DB 논문을 중심으로

박성택

(재)천안과학산업진흥원 전략기획본부장

A Study on Renewable Energy Research Trends Using Text Mining: Focusing on SCOPUS DB Papers

Seong-Taek Park

Division Director, CISTEP

요약 전세계적으로 기후 변화 문제가 심화되고 있으며, 화석연료의 사용 증가로 인한 다양한 문제들이 야기되고 있다. 특히 탄소중립 정책이 가속화됨으로 재생에너지에 대한 관심이 증가하고 있는 추세이다. 이에 본 연구에서는 SCOPUS DB를 활용하여 신재생에너지 연구 동향을 파악하였다. 초록 제공이 가능한 1,353개의 데이터를 확보하고 이를 분석할 수 있도록 데이터 전처리를 수행하고 분석을 수행하였다. 토픽모델링 분석 결과 중요한 키워드로 renewable와 energy로 나타났다. 이 외에도 electricity, solar, wind, 등이 중요한 키워드로 분석이 되었다. 본 연구 결과를 통해 신재생에너지 관련 기업의 실무자들이 신재생에너지 관련 연구동향을 실무에 활용할 수 있을 것으로 기대한다.

주제어 : 에너지, 재생에너지, 텍스트마이닝, 연구동향, SCOPUS

Abstract The problem of climate change is intensifying around the world, and the increasing use of fossil fuels is causing a variety of problems. In particular, interest in renewable energy is increasing due to the acceleration of carbon neutrality policies. In this study, we utilized SCOPUS DB to identify trends in renewable energy research. We obtained 1,353 data for which abstracts were available and performed data preprocessing to analyze them. Topic modeling analysis showed that renewable and energy were the most important keywords. In addition, electricity, solar, wind, etc. were analyzed as important keywords. Through the results of this study, it is expected that practitioners of renewable energy-related companies can utilize the research trends related to renewable energy in practice.

Key Words : Energy, Renewable Energy, Textmining, Research Trends, SCOPUS

1. 서론

기술의 발전으로 인해 많은 변화가 일어나고 있다. 특히 인간의 활동에 의한 온실가스의 배출로 인해 지구가 점점 더워지고 있는 실정이다. 탄소배출량을 신속 및 과감하게 줄이지 않는다면, 2040년경에는 1.5도 이상의

기온상승이 넘어갈 것으로 보인다.

예를 들면, 지구의 온도가 1.5도 상승을 한다면 폭염이 나타날 확률이 2배 증가하고 2도 상승한다면 폭염이 나타날 확률이 4배 증가를 하게 된다. 또한 1.5도 때보다도 2도가 올라갈 경우, 작물의 수확량이 감소하여 식량의 감소도 2배 이상 늘어날 것으로 전망된다[1].

*Corresponding Author : Seong Taek, Park(solpherd@gmail.com)

Received July 25, 2023

Accepted September 21, 2023

Revised August 7, 2023

Published September 30, 2023

최근 EU 집행위원회는 우크라이나 전쟁으로 촉발되고 있는 에너지 안보의 불안에서 벗어나기 위해서 2030년 재생 에너지 목표를 기존 40%에서 45% 수정하여 목표를 변경하였다[2].

정부에서도 탄소중립 2050에 발맞추어 탄소중립녹색성장 위원회를 발족하였으며, “기후위기 대응을 위한 탄소중립·녹색성장 기본법”, “2050 탄소중립위원회의 설치 및 운영에 관한 규정”을 제정을 하였다[3].

이를 통해 기후위기 대응에 동참을 하고, 국제 경제 질서와 급변하는 무역환경의 변화에 선제적인 대응을 하고 있다. 현재 EU·미국 탄소국경조정제도 도입계획 발표하였으며, 글로벌 기업의 RE100 선언, ESG 투자 확산 등에 선제적 대응 필요하다고 할 수 있다.

탄소중립에서는 재생에너지가 핵심이지만, 이에 대한 관련 연구는 아직까지 많이 부족한 실정이다. Scopus 데이터베이스에서 re100을 검색하면, 관련 연구 논문이 부족한 실정이다. 특히 국내에서는 특정한 분야의 기술 중심으로만 연구가 진행되고 있는 바 본 연구에서는 SCOPUS DB의 논문을 통해 재생에너지(Renewable Electricity) 논문의 연구 동향을 살펴보고 이를 통해 시사점을 도출하고자 한다.

2. 이론적 배경

2.1 재생에너지

재생 에너지는 재생이 가능한 에너지로서, 햇빛, 물, 지열, 강수, 생물 유기체 등에 포함된 에너지를 변환시켜 이용하는 에너지를 의미한다. 이러한 재생 에너지에는 태양 에너지, 풍력, 수력, 해양 에너지, 지열 에너지, 바이오 에너지가 있다.

재생 가능 에너지의 종류는 여러 가지가 있지만, 이것들 대부분은 태양 에너지의 변형이기 때문에 우리가 하루에 사용할 수 있는 재생 가능 에너지의 양은 하루 동안 지구로 들어오는 태양 에너지의 양을 넘지 못하기 때문에 우리가 무한한 에너지를 얻을 수 있는 것은 아니다.

그러나 재생 에너지는 일단 기술력이 갖추어지게 된다면 매우 경쟁적으로 전기를 생산할 수 있고, 기후 변화 문제와 화석 연료의 고갈 문제를 해결할 수 있기 때문에 반드시 필요하고 개발되어야 할 기술이다[4].

재생에너지 보급의 각국의 정책을 살펴보면, 우선 전력 부문에서는 재생에너지 보급을 위한 인센티브와 의

무화를 부여한 국가는 지속적으로 증가하고 있지만, 냉난방부문에 대한 정책적 관심은 상당히 저조한 편으로 나타났다[5].

2.2 텍스트마이닝

텍스트 마이닝이란 인간(사람)의 언어로 이루어진 비정형 데이터들 중에서 자연어처리기법(NLP)을 활용하여 대량의 텍스트 데이터 중에서 정보를 추출하고 연관성을 분석하고, 분류 및 군집화 등을 통해 텍스트 데이터 속에서 숨겨진 의미를 발견하는 기법을 의미한다[6].

텍스트 마이닝은 비정형 데이터마이닝 기법 중에서 대표적인 기법으로 연구동향 분석에서 가장 많이 사용되고 있는 방법중에 하나이다[7,8].

2.3 LDA(Latent Dirichlet Allocation)

LDA는 잠재 디리클레 할당의 약자로, 주어진 문서에 대해 각 문서별로 어떠한 주제들이 존재를 하고 있는지를 나타내는 확률적 토픽 모델의 기법 중 하나이다[9].

토픽모델링은 텍스트의 본문에 숨겨져 있는 의미나 구조를 발견하기 위해서 사용되는 텍스트 마이닝 기법 중의 하나이다[10]. LDA 모델은 Fig.1과 같다.

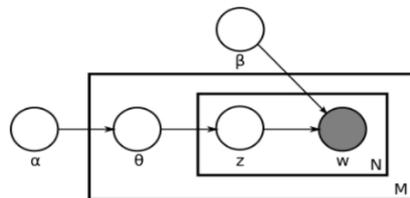


Fig. 1. LDA Model

기본적으로 LDA 가정은 문서의 생성과정이 말뭉치 전체의 문서 개수와 토픽수, 관찰가능한 변수 등의 다양한 정보를 가지고 하이퍼파라미터를 제외한 나머지 모든 잠재변수를 추정하는 과정을 수행한다.

또한 LDA는 기존 LSI 방법의 단점인 과적합화, 데이터 증가에 따라 모델의 매개변수가 증가하는 현상을 해결하고 있기 때문에 많이 사용이 되고 있는 방법론이다[11].

LDA는 학습 데이터가 없어도 문서에 대한 분류 작업 수행을 한다는 장점을 가지고 있기 때문에 본 연구에서는 LDA를 활용하여 분석을 수행하였다[12].

3. 연구방법 및 절차

3.1 연구방법

본 연구에서는 SCOPUS DB를 활용하여 학술논문을 검색하였다[12]. Title은 “Renewable Electricity”로 검색어를 한정하였다. 초록이 없는 데이터는 검색에서 제외를 하였으며, 2013년부터 2022년까지 10년간의 데이터를 활용하였다.

3.2 연구절차

본 연구에서는 Fig. 2에서 보는 바와 같이 가지 절차를 통해 분석을 수행하고자 한다. 먼저 SCOPUS DB에서 논문 초록 데이터를 수집한다. 수집된 데이터는 정제(전처리) 과정을 거쳐서 분석에 적합하도록 하였다.



Fig. 2. Analytical Procedure

데이터 분석은 R을 활용하여 수행하였으며, TD-IDF, LDA 분석 기법을 사용하여 분석을 수행하였다. 도출된 결과를 시각화하였다[11,13].

4. 분석결과

4.1 자료분석

본 연구에서 분석에 사용한 저널은 총 1,353편이며, Fig. 3에서 보는 바와 같이 재생에너지 논문은 2017년부터 증가하기 시작하였으며, 2022년 5월 현재 90편으로 나타났다.

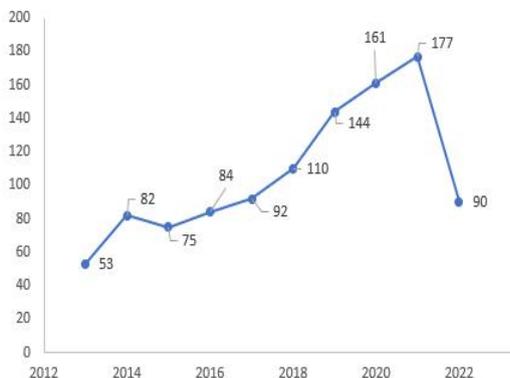


Fig. 3. Number of papers

Subject Area는 Table 1에서 보는 바와 같다. Energy가 970편으로 가장 많이 분포를 하고 있으며, 그 다음으로는 Environmental Science, Engineering, Economics, Econometrics and Finance 순이다.

Table 1. Subject Area

Subject Area	Paper
Energy	970
Environmental Science	548
Engineering	425
Economics, Econometrics and Finance	156
Social Sciences	150
Mathematics	115
Business, Management and Accounting	93
Computer Science	69
Chemical Engineering	43
Materials Science	35

Table 2에서 보는 바와 같이 키워드별로 살펴보면 Electricity Generation이 420편으로 가장 많고, 그 다음으로는 Alternative Energy, Renewable Energy Resources, Renewable Resource, Renewable Energy 순으로 나타났다.

Table 2. Keyword

Keyword	Paper
Electricity Generation	420
Alternative Energy	382
Renewable Energy Resources	381
Renewable Resource	368
Renewable Energy	351

Fig. 4에서 보는 바와 같이 국가별로 살펴보면, 미국이 229편으로 가장 많은 논문을 게재하고 있다. 그 다음으로는 독일, 중국, 영국, 스페인 순이다

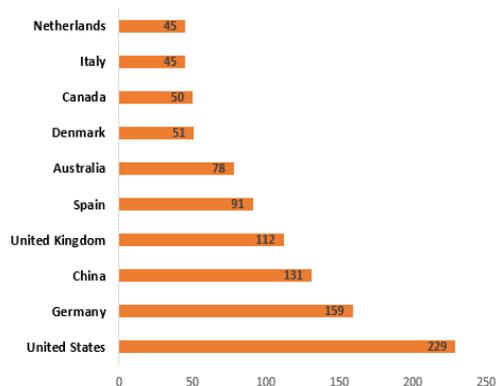


Fig. 4. Number of papers by country

4.2 분석결과

토픽 분석 결과는 다음과 같다. LDA 그래프에서 살펴보면, 오른쪽은 전체 코퍼스(corpus)에서 가장 중요한 키워드가 바(bar) 차트 형태로 나타난다. 왼쪽 사분면에는 5개 토픽이 파이 차트 형태로 나타난다[14,15].

Fig. 5에서 보는 바와 같이 1번 토픽에서 중요 키워드로는 renewable, electricity, solar, wind,

generation, system 등이 있으며, 토픽의 주제는 'renewable' 임을 유추할 수 있다.

Fig. 6에서 보는 바와 같이 2번 토픽에서 중요 키워드로는 electricity, energy, renewable, system, power, model 등이 있으며, 토픽의 주제는 'electricity' 임을 유추할 수 있다.

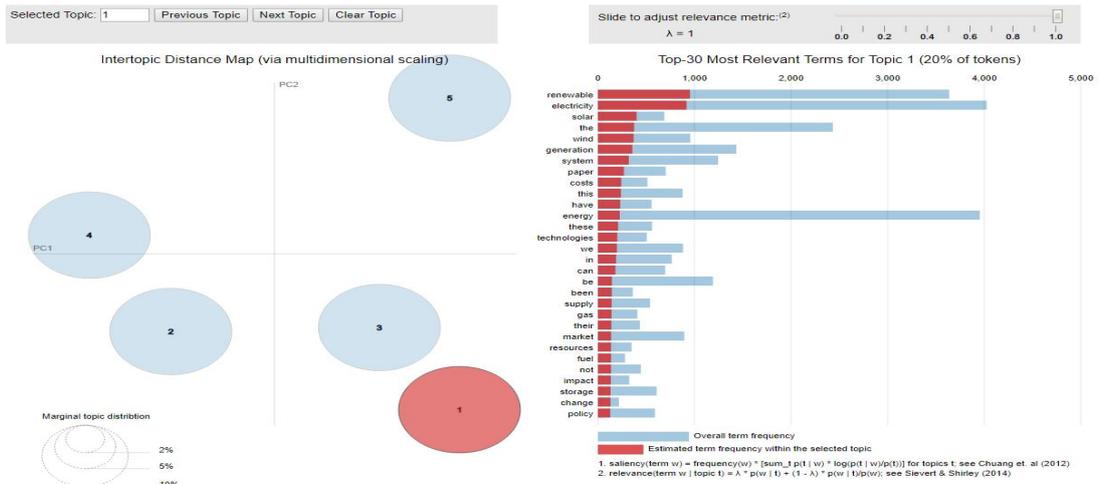


Fig. 5. LDA analysis results 1

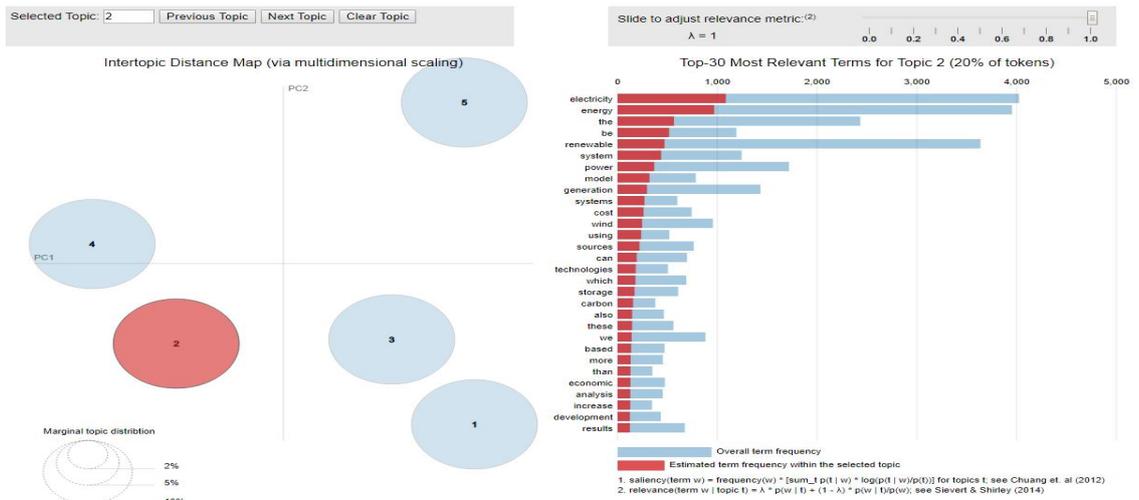


Fig. 6. LDA analysis results 2

Fig. 7에서 보는 바와 같이 3번 토픽에서 중요 키워드는 renewable, electricity, energy, study, market, cost, generation 등이 있으며, 토픽의 주제는 'renewable' 임을 유추할 수 있다.

Fig. 8에서 보는 바와 같이 4번 토픽에서 중요 키워드로는 energy, electricity, power, generation,

sources, demand, results 등이 있으며, 토픽의 주제는 'energy' 임을 유추할 수 있다.

Fig. 9에서 보는 바와 같이 5번 토픽에서 중요 키워드로는 energy, electricity, power, generation, sources, demand, results 등이 있으며, 토픽의 주제는 'energy' 임을 유추할 수 있다.

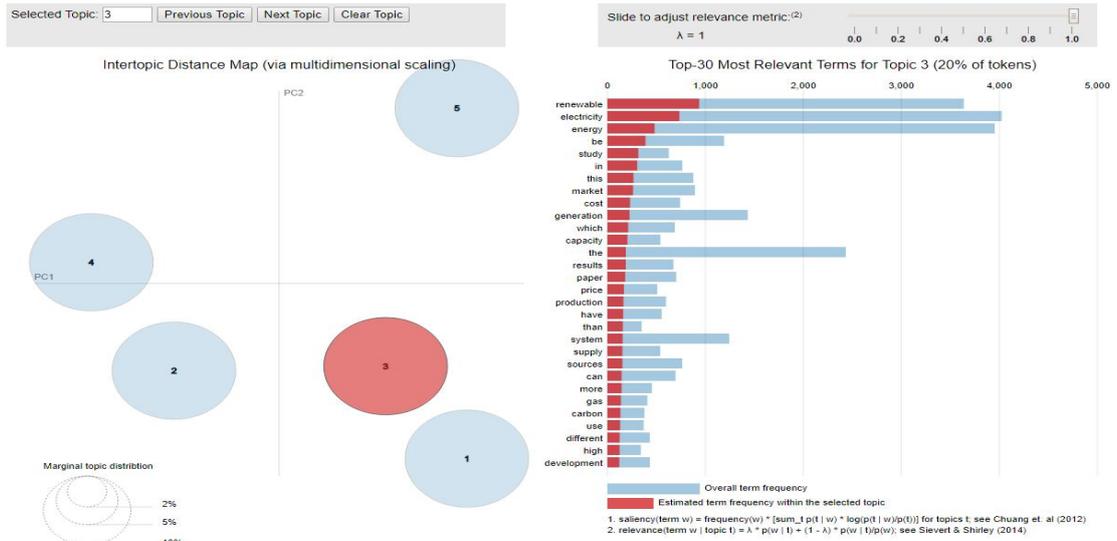


Fig. 7. LDA analysis results 3

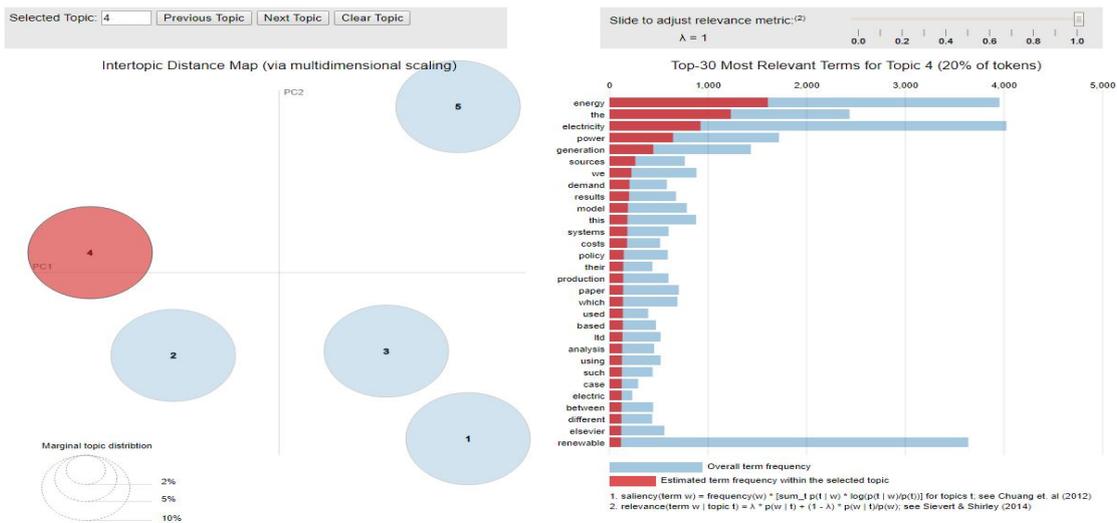


Fig. 8. LDA analysis results 4

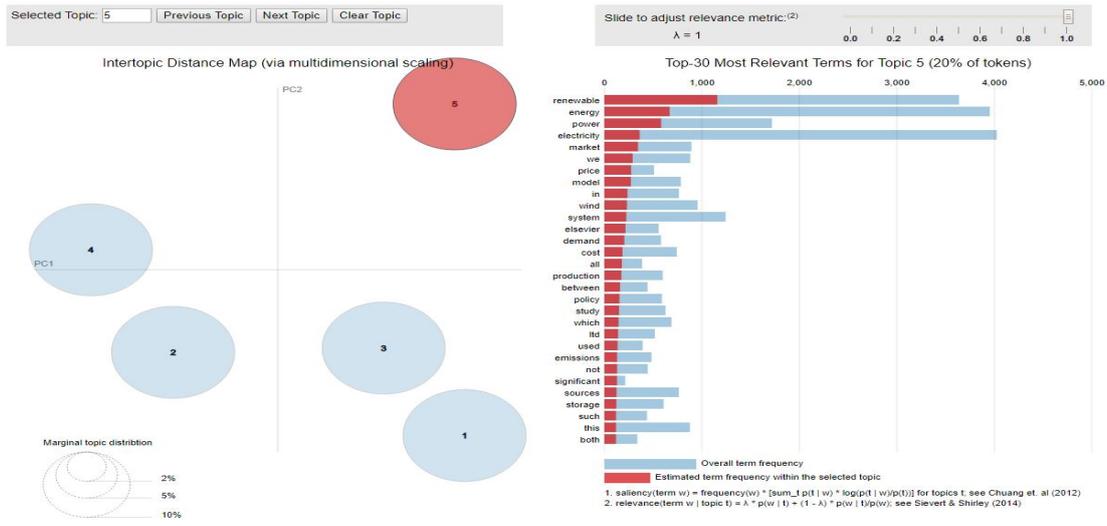


Fig. 9. LDA analysis results 5

4.3 논의

토픽모델링 분석 결과 중요한 키워드로 renewable 와 energy로 나타났다. 이 외에도 electricity, solar, wind, 등이 중요한 키워드로 분석이 되었다.

최근들어 신재생에너지에 대한 국가와 기업들의 관심이 점차 고조가 되고 있다. 특히 탄소중립 2050 선언 이후 과거와는 다르게 재생에너지에 대한 관심과 연구가 지속적으로 증가하고 있다. 특히 태양, 바람 등 신재생에너지에 대한 중요성이 점차 증가하는 것으로 나타났다.

5. 결론

본 연구는 재생에너지 연구동향을 파악하고자 SCOPUS DB에서 Renewable Electricity라는 주제어(keyword) 만을 검색하여 데이터를 추출하였으며, 논문의 초록(abstract) 수집이 가능한 총 1,353개의 데이터를 수집하고 이를 전처리(pre-processing)를 거쳐 분석을 수행하였다. 분석결과는 다음과 같다.

재생에너지 논문은 2017년 부터 증가하기 시작하였으며, 2022년 5월 현재 90편으로 나타났다. 토픽 모델링 분석 결과, renewable, electricity, energy가 중요한 토픽으로 도출되었다.

본 연구에서는 기존에서는 아직 고려하지 않았던 재생에너지의 연구 동향을 파악하고자 SCOPUS 데이터베이스를 사용하고 TD-IDF, LDA, 토픽 모델링 방법을

활용하여 분석을 시도하였다는 점에서 학술적인 의의가 있다고 하겠다.

이를 통해 향후 재생에너지 관련 연구 동향과 발전 가능성을 판단하는데 있어 중요한 의의가 있다. 또한 본 연구의 결과는 재생에너지 관련 기업의 실무자들에게 재생에너지 연구에 대한 동향을 파악할 수 있게 하여 실무에서도 활용가능할 것으로 보인다.

그러나 본 연구에서 SCOPUS DB로 한정을 하여 연구 동향을 파악하였기 때문에 일반화할 수 없다는 점과 국내의 재생에너지 연구 동향을 파악할 수 없었다는 점과 정량적인 분석만을 수행하였다는 점에서 한계점을 가지고 있다.

향후 연구에서는 다양한 국내외 학술 데이터베이스를 활용하는 것이 필요해 보이며, 국내와 해외 논문의 비교, 국가간의 비교 등 추가적인 분석이 필요할 것으로 보인다.

REFERENCES

[1] J. Marotzke & P. M. Forster. (2015). Forcing, feedback and internal variability in global temperature trends. *Nature*, 517(7536), 565-570.
 [2] <https://www.hani.co.kr/arti/society/environment/1043462.html>
 [3] <https://www.2050cnc.go.kr>
 [4] <https://if-blog.tistory.com/5481>
 [5] J. Y. Kong & I. H. Cho. (2021). *Analysis of*

International New and Renewable Energy Policy and Market. Basic Research 21-27. Korean Energy Economics Institute.

- [6] E. M. Park (2018). *Effect of Technology Innovation Competency and Patent Competency on the Business Performance*. Master's thesis. Kyungpook National University. Daegu.
- [7] S. T. Park, D. Y. Kim, & G. Li. (2021). An analysis of environmental big data through the establishment of emotional classification system model based on machine learning: focus on multimedia contents for portal applications. *Multimedia Tools and Applications*, 80, 34459-34477.
- [8] M. H. Ko, E. M. Park, & S. T. Park. (2018). Mining the open science: Themes and trends 10 years of open science. *Journal of Advanced Research in Dynamical and Control Systems*, 10(1), 263-270.
- [9] S. T. Park, & Y. K. Kim. (2020). Artificial Intelligence Research Trend Using SCOPUS DB: 2011-2019. *International Journal of Advanced Science and Technology*, 29, 304-312.
- [10] S. T. Park & C. Liu, (2022). A study on topic models using LDA and Word2Vec in travel route recommendation: focus on convergence travel and tours reviews. *Personal and Ubiquitous Computing*, 26, 429-445.
- [11] S. C. Deerwester, S. T. Dumais & T. K. Landauer. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6), 391-407.
DOI : 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9
- [12] E. M. Park & J. H. Seo. (2020). Analysis of Research Trends in Technology Innovation: Focus on SCOPUS DB. *Journal of Convergence for Information Technology*, 10(8), 120-126.
- [13] S. T. Park, S. W. Lee & T. G. Kang. (2018). A study on the trend of cloud service and security through text mining technique. *International Journal of Engineering & Technology*, 7(2.33), 127-132.
- [14] E. M. Park & J. H. Seo. (2020). Analysis of Research Trends in Technology Innovation: Focus on SCOPUS DB. *Journal of Convergence for Information Technology*, 10(8), 120-126.
- [15] I. G. Kim et al. (2021). Analysis on Research Trends in Sport Facilities: Focusing on SCOPUS DB. *Journal of Industrial Convergence*, 19(6), 11-19.

박 성 택(Seong Taek Park)

[중신회원]



- 2010년 2월 : 충북대학교 경영정보학과(경영학박사)
- 2011년 7월 ~ 2012년 6월 : 성균관대학교 박사후연구원
- 2020년 3월 ~ 2021년 5월 : 선문대학교 SW융합학부 SW융합경영전공 교수
- 2021년 6월 ~ 현재: (재)천안과학산업진흥원 전략기획본부장
- 관심분야 : 빅데이터, 인공지능, 특허, 특허가치평가, MIS 등
- E-Mail : solpherd@cistep.re.kr