

# 신재생에너지 동향 파악을 위한 토픽 모형 분석

신규식<sup>1</sup>, 최회련<sup>1</sup>, 이홍철<sup>\*</sup>

<sup>1</sup>고려대학교 산업경영공학과

## Topic Model Analysis of Research Trend on Renewable Energy

KyuSik Shin<sup>1</sup>, HoeRyeon Choi<sup>1</sup>, HongChul Lee<sup>\*</sup>

<sup>1</sup>Department of Industrial and Management Engineering, Korea University

**요약** 기후변화 및 환경오염에 대응하기 위해 신재생에너지 정책 연구가 증가하고 있다. 신재생에너지는 녹색산업과 녹색 기술로 대표되는 새로운 성장 동력 기술이다. 현재 우리나라는 태양광, 풍력, 수소연료전지 등 3대 전략부분에 신재생에너지 보급 및 기술개발사업에 대한 투자가 이루어지고 있지만 아직은 초기 단계로, 연구 방향 및 투자 분야에 대한 불확실성을 줄이는 것이 무엇보다도 시급한 실정이다. 따라서 본 연구는 빅데이터(big data) 분석방법 중 텍스트 마이닝(Text mining method)과 토픽 모델링 기법(multinomial topic model)을 신재생에너지와 관련된 최근 10년간의 우리나라 언론기사에 적용하여 국가 정책의 핵심이슈 및 세계적인 연구 트렌드를 분석하고, 성장 가능성이 있는 신재생에너지 분야를 예측하였다. 정보통신기술을 바탕으로 한 연구결과는 신재생에너지 분야에 활발히 적용될 것으로 예측된다.

**Abstract** To respond the climate change and environmental pollution, the studies on renewable energy policies are increasing. The renewable energy is a new growth engine technology represented by the green industry and green technology. At present, the investments for the renewable energy supply and technology development projects of three main strategy sectors such as sunlight, wind power and hydrogen fuel cell are implemented in our country, while they are still in the early stage, accordingly reducing those uncertainty for the research direction and investment fields is the most urgent issue among others. Thus, this study applied text mining method and multinomial topic model among the big data analysis methods on our country's newspaper articles concerning the renewable energy over the last 10 years, and then analyzed the core issues and global research trend, forecasting the renewable energy fields with the growth potential. It is predicted that these results of the study based on information and communication technology will be actively applied on the renewable energy fields.

**Keywords** : Big Data, Multinomial topic model, News article, Renewable energy, Text analysis

### 1. 서론

지속적인 산업의 발전과 기후변화로 인한 지구온난화, 자원 전쟁 등과 같은 산업과 사회에 미치는 부정적인 효과로 인하여 환경에 대한 관심과 연구가 증가하고 있다. 이와 관련된 대표적인 분야인 녹색성장을 위한 신재생에너지의 개발 및 보급을 포함한 그린정책에 관련된

연구가 급증하고 있다. 신재생에너지는 녹색산업과 녹색 기술로 대표되는 새로운 성장 동력 기술이며, 에너지 분야의 공급과 수요 관리기술 부분 등에 대한 투자는 신재생에너지의 보급·확산뿐만 아니라 에너지 고효율화와 청정에너지 공급확대, 정보통신기술(ICT) 기반의 새로운 산업 창출 및 마이크로 그리드 전원 제어 시스템과 산업계 배출권 거래제대응에 대한 기술개발을 발생시켰다.

본 논문은 교육부 및 한국연구재단의 BK21 플러스 사업(고려대학교 제조·물류 분야에서의 빅데이터 운용 사업팀)으로 지원된 연구임.

\*Corresponding Author : Hong-Chul Lee(Korea Univ.)

Tel: +82-2-3290-3267 email: hclee@korea.ac.kr

Received August 18, 2015

Revised (1st September 2, 2015, 2nd September 9, 2015)

Accepted September 11, 2015

Published September 30, 2015

뿐만 아니라 기존의 석탄, 석유를 이용한 화석연료가 아닌 에너지 고갈의 위험이 없는 친환경적인 신재생에너지 산업이 주목을 받고 대체에너지, 신에너지 개발, 태양에너지, 탄소제로산업 등의 연구와 언론 정보, 특허 등이 지속적으로 증가하고 있으며 기업들도 신재생에너지와 관련된 외국 업체와의 협력과 환경 문제에 관련된 사업이 추진되는 등 에너지 환경에 대한 인식의 변화가 생겼다.

현재 우리나라는 태양광, 풍력, 수소연료전지 등 3대 전략부문에 신재생에너지 보급 및 기술개발사업에 대한 투자가 이루어지고 있지만 아직은 초기 단계로 재생에너지 기술의 세계적인 연구 트렌드, 국가정책 방향 등을 고려한 연구 및 투자는 부족한 실정으로 연구방향 및 투자 분야에 대한 불확실성을 감소시켜 줄 수 있는 방법론이 필요하다.

따라서 본 연구에서는 신재생에너지와 관련된 내용을 담고 있는 지난 10년간의 언론 기사에서 신재생에너지의 핵심 키워드들을 분류하여 국가정책 및 세계적 연구 트렌드를 기반으로 정책적, 사회적, 연구적, 그리고 기술적 동향을 동시에 파악할 수 있는 분석모델을 제시한다. 제시하는 분석모델은 분석데이터가 빅데이터(big data)이므로, 첫 단계로 분석의 용이성을 위한 데이터 전처리 과정과 두 번째는 빅데이터 분석 기술 중의 하나인 토픽 모델링 기법(multinomial topic model)을 적용하고, 마지막으로 분석 결과간의 유기적인 연관성을 그래프로 표시하는 기능들로 구성되어 있다. 본 연구에서 결과로 제시되는 트렌드와 국가정책 등이 반영된 신재생에너지의 동향은 에너지 분야의 연구자들에게 기초 연구 자료로서의 활용뿐만 아니라 앞으로의 연구 방향에 도움을 줄 수 있을 것으로 기대된다.

## 2. 선행연구

그동안 신재생에너지 연구는 경제성 분석, 정책적 비교 및 타당성 분석, 에너지 시장평가 및 기술 분석이 진행되었으며 이는 효율적인 신재생에너지 활용 방안에 초점이 맞추어져 있다. 한편, 신재생에너지 사업에 대한 투자 방안과 미래의 연구주제, 타 분야와의 융합연구를 위한 기술예측 및 개발 관련 연구는 미비한 실정이다. 따라서 지금까지의 신재생에너지 연구 주제와 내용을 가지고, 앞으로의 신재생에너지의 기술예측과 연구동향을 파악할 필요가 있다.

이를 위한 방법론으로는 다양한 분야의 수많은 문헌들 사이에서 분야와 관련된 주제를 정리, 검색 및 이해할 수 있도록 해주는 토픽 모델링이 있다.

토픽 모델링은 (1) 많은 자료들 중에서 특정 주제를 찾아내고, (2) 찾아낸 주제를 바탕으로 문헌들에 특정 주석을 달아주며, (3) 문헌에 달린 주석을 기반으로 문서를 예측하여 정리하고 요약해주는 과정을 가지는 기법으로 문헌의 주제를 찾아내는데 사용된다. 최근에는 문헌에서 확률모형을 바탕으로 문헌의 주제를 찾아내는 LDA(Latent Dirichlet Allocation)가 많이 사용되고 있다. 토픽모델링은 문헌 주제들의 집합이고, 문헌은 단어들로 구성된다고 가정할 때 문헌 별 주제나 단어의 출현 확률을 바탕으로 각 주제들을 파악할 수 있게 된다[1]. 또한 데이터의 최적 표현의 관점에서 데이터의 축소 방법으로 문서별로 발생확률이 높은 토픽간의 관계를 바탕으로 토픽간의 네트워크를 제시해주고 있다.

기존의 LDA 기법을 통해 문서들의 순서를 고려하여 시간의 흐름에 따른 토픽의 변화를 추적하는 연구를 수행하였다[2]. LDA 기법은 텍스트로 이루어진 문서 컬렉션에서 적용할 수 있을 뿐만 아니라, 이미지, 유전 정보, 통계수치 등 여러 데이터에서도 패턴을 추출할 수 있다. 이러한 토픽모델링은 문헌정보학, 건축학, 산업공학, 지식재산 등의 다양한 분야에서 활용되고 있다.

진철아[3]는 트위터 데이터를 분석하여 토픽의 변화 시점 및 패턴을 파악하여 트위터는 언론 뉴스에 즉각적으로 반응하며 부정적 이슈를 빠르게 확산시키는 것을 확인하였다. 박자현[4]은 국내 문헌 정보학의 연구 동향을 확인하기 위해 한국 비블리아 학회지의 발표 논문을 수집하여 LDA기반의 토픽모델링 실험을 수행하여 문헌 정보학자들의 많은 연구 주제가 발견된 학문을 확인하고, 주요 연구주제의 연도별 추이를 분석하여 연구 동향을 파악하였다.

Griffiths[5]는 미국국립과학원회보(Proceedings of the National Academy of Sciences of the United States of America)의 논문을 대상으로 토픽모델링 기법을 적용하여 의미 있는 토픽을 발견함과 동시에 최근 관심이 상승세인 연구 주제(Hot Topic)와 관심이 하강세인 연구 주제(Cold Topic)를 나타냈다.

연구동향 파악에 토픽모델링이 유용하게 적용된다고 했다. 이외에도 많은 연구 들은 토픽모델링을 통해 다양한 텍스트 자원을 분석하고 수행하였다. 선행 연구들은 토픽모델링 기법을 방대한 문서에서 단어를 클러스터링

하는 방식으로 주제를 도출한다.

따라서 본 연구는 토픽모델링기법을 이용하여 신재생 에너지 관련 언론기사의 주제와 내용을 도출하고, 시간적 흐름에 따른 변화추이를 분석하여 연구동향을 파악하고자 한다.

### 3. 연구방법

본 연구의 방법은 데이터 수집, 데이터 전처리 과정인 형태소분석, 토픽모델링 방법의 절차를 통해 신재생에너지 연구동향을 분석한다.

#### 3.1 데이터 수집

신재생에너지 산업의 키워드 및 분석은 특히, 연구보고서, 연구논문, 언론기사 등 다양한 텍스트 문서로 도출되고 있다. 이러한 문서 중 언론기사는 국가 정책, 세계 연구 트렌드가 표출되고 있어 정책적, 사회적, 기술적 동향을 파악하기 적합하다고 판단된다.

본 연구에서는 최근 10년간, 2006년 1월부터 2015년 6월까지의 신재생에너지 관련 언론기사를 대상으로 데이터를 수집하고 분석한다. 신재생에너지 관련 언론기사의 데이터는 자바스크립트의 크롤러(Crawler)를 적용하여 데이터를 수집하였다. 연구 데이터는 핵심키워드가 포함된 문장을 분석하였다. 뉴스 제목, 내용의 중복성을 피하기 위해 중복된 기사를 제거하였고, 내용의 단어들 이 70% 이상 중복된 정보는 연구 데이터에 포함시키지 않았다.

#### 3.2 형태소 분석

자연어 처리는 언어 현상을 기계적으로 분석하고 컴퓨터가 이해 할 수 있는 형태로 변경하고 다시 인간이 이해할 수 있는 언어로 표현하는 제반 기술을 의미한다. 뉴스 기사 정보는 비정형 텍스트 문서로 텍스트의 전처리 과정이 필요하다. 정보 추출이 가능한 형태로 변환이 요구된다. R Project의 KoNLP 패키지를 사용한다. KoNLP는 한글 형태의 단어를 형태소 단위로 인식할 수 있는 기능을 제공하고 있다. 두 글자 이상으로 된 명사형태의 단어를 추출한다. 형태소 분석은 수집된 정보를 기계적으로 분석하고 토픽 도출을 위한 형태소 분석이 선행되어야 한다. 형태소 분석은 미등록어, 오타자, 띄어쓰기 오류 등을 해결하기 위해 자동 색인 방지 목적으로 불용어를 파악해야 한다.

#### 3.3 토픽 모델링

문서에서 주제가 되는 키워드를 추출하기 위해 LDA(Latent Dirichlet Allocation) 기반 토픽모델링 기법을 사용하였다. LDA는 주성분분석(Primary Components Analysis)과 유사하게 데이터의 차원을 축소하는데 유용하다. 일관성을 가진 의미 있는 주제를 클래스 집단으로 생성해 주는 장점으로 비정형 텍스트 분석에 사용되고 있다[6].

LDA는 Dirichlet 분포로 주어진 문서에서 잠재적으로 갖는 주제들을 추론해 내는 확률모델로 Generative topic model이다[7-9]. 문서 정보에 대한 주제들의 분포  $\theta$ 와 각 주제별로 특정 단어 생성될 확률 분포  $\beta$ 가 주어졌을 때, 특정 문서가 만들어질 확률을 선택하고 선택된 주제에 대해 존재하는 키워드들을 확률적으로 선택하는 과정을 반복함으로써 임의의 문서를 생성한다. 주어진 문서에 대한 파라미터들을 통해 문서의 앞에서부터 단어를 채울 때 마다 하나의 주제를 선택하고 다시 그 주제로부터 단어를 선택하는 확률적 방식으로 주제를 추론하는 모델이다[Fig. 1].

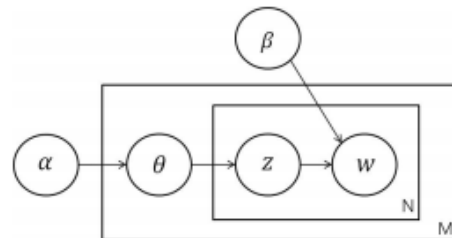


Fig. 1. Latent Dirichlet allocation Model

LDA는 모델링을 수행하기 위한 샘플링의 반복 횟수와 추출할 주제의 수를 지정해야 한다. 일반적인 문서에서 파라미터 추정하는 기법은 깁스샘플링(Gibbs Sampling) 알고리즘을 이용한다. 샘플링을 기반으로 추론하는 방법으로 널리 사용되고 있는 MCMC(Markov Chain Monte Carlo) 방법으로 두 개 이상의 변수들로부터 연쇄적으로 표본을 샘플링 하는 것으로 분포를 추정하는 깁스샘플링은 주어진 모수에 의해 자료를 분류하고 분류된 자료 속에서 모수를 다시 추정한다.

본 연구에서 깁스 샘플링을 2000회 반복하여 총 20개의 토픽 주제를 추출하였다. 추출된 주제에 대한 적절한 주제명을 선정하고 신재생에너지 산업의 트렌트에 대해 분석하였다.

### 4. 연구 결과

본 연구에서는 신재생에너지 연구 동향을 분석하는데 목적이 있기에, 전처리를 통해 일정한 키워드가 포함된 언론기사의 제목 내용을 비교하고, 중복되지 않은 항목을 도출하였다. 최근 10년간 신재생에너지 관련 언론기사는 총 51,558개로, 2006년부터 현재까지 꾸준히 증가하고 있음을 알 수 있다[Fig 2].

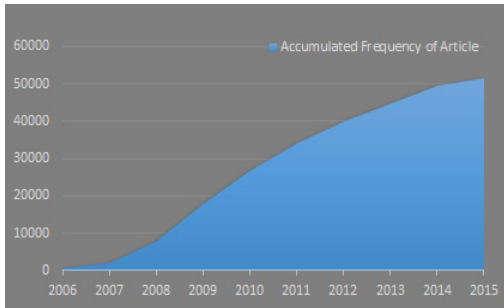


Fig. 2. An Analysis result of annual accumulated frequency of article( or news)

수집된 데이터를 명사형태의 단어들로 추출한 다음, 3,000회 이상 인용된 단어의 빈도수를 통해 시각화하였다(Fig 3). 신재생에너지의 키워드로는 재생, 에너지, 산업, 녹색기술, 친환경, 성장 동력, 온실가스, 태양광, 풍력 발전, 연료전지가 나타났다. 이는 에너지산업, 에너지원, 환경문제, 미래계획 등에 대한 이슈가 있음을 확인할 수 있다.



Fig. 3. An Analysis result of WordCloud

토픽모델링 기법을 이용하여 신재생에너지의 연구 동향을 분석하였다. 언론 기사는 본문 내용의 형태소 분석

과 2글자 미만의 단어는 제외하고, 불용어 처리 및 Corpus 함수를 통해 하나의 집단을 만들고 51,558개의 문서에서의 188,248개의 용어로 이뤄진 Document Term Matrix의 구조화된 데이터를 구축하였다. 상위 20개의 토픽은 주제별 사후 출현확률이 높은 단어로 주제를 선정하고 20개의 토픽 주제를 도출할 수 있었다. 각 토픽들은 에너지협력사업, 에너지 절약 교육프로그램, 탄소배출권, 선진국의 온실가스 감축 관련 노력 현황, 스마트 그리드, Green Home, 신재생에너지사업, 태양광보급사업, 폐자원을 활용한 에너지 현황, 자유무역협정, 지역 경제 활성화 사업, 국내 기업 사업 현황, 신재생에너지 투자현황, 태양전지 관련 사업, 에너지와 IT기술 접목 사업, 정부 주도 포럼, 신재생에너지육성사업, 국제 그린 에너지 엑스포, 에너지 인재 양성 프로그램, REC(Renewable Energy Certificate)로 분류 결과를 나타낼 수 있다.

분류한 주제토픽들은 기업들의 해외진출, 스마트 그리드 관련된 AMI, Green Home, 에너지 효율을 향상을 위한 저탄소 기술 및 지능형 전력계량 인프라, 환경 문제에 대한 온실가스 감축 등과 연관 지어 볼 수 있다.

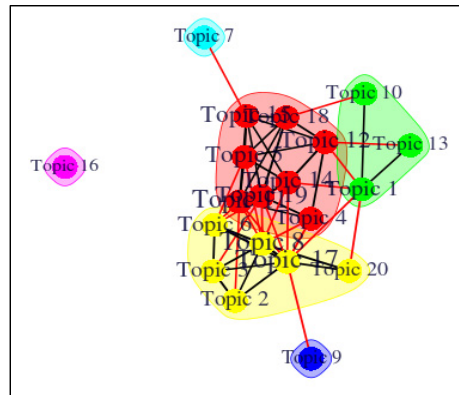


Fig. 4. Community Detection Algorithms

많은 데이터를 가진 네트워크 모델은 육안으로 특정한 관계를 식별하기 힘들다. 토픽모델링의 네트워크 모델은 토픽 개체의 변수들의 위치 중심성 구조로 유사성을 확인할 수 있는 Community detection 알고리즘을 통해 확인이 가능하다[10-12]. 분석 결과로 20개의 주제들은 6개의 집단으로 Walktrap 알고리즘을 통해 무작위 기반의 네트워크로 확인하였다[Fig. 4].

신재생에너지사업, 폐기물에너지 사업, 기업의 해외

Table 1. Koreans Analysis result of TOPIC modeling

	Topic Word
Topic 1	에너지, MOU, 태양광, 양해각서, 프로젝트, 풍력발전, 태양광발전, 대규모, 에너지사업, UAE, 저수지, 업무협약, 농어촌, 소수력, 말레이시아, 한국전력, 바이오, 건설사업, 에너지개발, 마스다르
Topic 2	에너지, 자전거, 친환경, 일자리, 프로그램, 활성화, 전시관, 어린이, 청소년, 홍보관, 지역경제, 기후변화, 저탄소, 창원시, 조성사업, 만들기, 관광자원, 스포츠, 박물관, 경기장
Topic 3	에너지, 온실가스, 기후변화, 저탄소, 에너지절약, 이산화탄소, 배출량, CDM, 친환경, 에너지효율 기후변화협약, 화석연료, 활성화, 효율화, 에너지관리, 지구온난화, 고효율, Green, 개발체제, 프로그램
Topic 4	에너지, 태양광, 원자력, 화석연료, 우리나라, 에너지원, 선진국, 온실가스, 기본계획, 에너지산업, 경쟁력, 가능성, 화석에너지, 보급률, 원자력발전, 태양열, 연료전지, 성장동력, 고유가, 효율성
Topic 5	에너지, 스마트, 그리드, 저장장치, 시스템, 태양광, 자동차, 지능형, 배터리, 솔루션, 인프라, 마이크로, 차전지, 차세대, 친환경, 비즈니스, 실시간, 하이브리드, 성장동력, 효율화
Topic 6	에너지, 시스템, 친환경, 태양광, 태양열, 건축물, 냉난방, 아파트, 서울시, 고효율, 사용량, 저탄소, 에너지효율, 신도시, 태양광발전, 하우스, 디자인, 절약형, 공동주택, 최소화
Topic 7	신재생에너지, 부동산, 서비스, 컴퓨터, 대체에너지, 자동차, 컨설팅, 판매업, 광역시, 에너지사업, 소프트웨어, 기계설비, 제조업, 인터넷, 제조판매, 공사업, 엔지니어링, 원자재, 전기전자, 글로벌
Topic 8	에너지, 태양광, 보급사업, 태양열, 에너지관리, 발전시설, 태양광발전, 지원사업, 발전설비, 공공기관, 저탄소, 보조금, 에너지원, 활성화, 연료전지, 시스템, 그린빌, 공공시설, 사회복지시설, 친환경
Topic 9	에너지, 바이오, 폐기물, 친환경, 재활용, 에너지사업, 쓰레기, 보일러, 폐자원, 플랜트, 하수처리장, 생활폐기물, 슬러지, 에너지원, 매립지, 에탄올, 온실가스, 지방행정, 열병합발전소, 부산물
Topic 10	에너지, 대통령, 비즈니스, 글로벌, 프로젝트, 원자력, 연구소, 인프라, ict, 협력방안, 덴마크, 양국간, 정상회담, 경제협력, 전문가, mou, 정보통신기술, 양해각서, 스페인, 자유무역협정
Topic 11	에너지, 새만금, 클러스터, 에너지산업, 전남도, 호남권, 친환경, 자동차, 경제권, 프로젝트, 전략산업, 태양광, 동북아, 광주시, 투자유치, 조성사업, 성장동력, 수도권, 세종시, 저탄소
Topic 12	에너지, 플랜트, 태양광, 성장동력, 자원개발, 에너지사업, 경쟁력, 글로벌, 현대중공업, 국내외, 칼텍스, 신사업, 다각화, 원자력, 풍력발전, 차세대, 기술력, 신규사업, 해외자원개발, 중공업
Topic 13	에너지, 관련주, 풍력발전, 에너지사업, 가능성, 자회사, 긍정적, 기대감, 경남신재생에너지, 계열사, 하반기, 애널리스트, 투자자, 상승세, 전자재료, 영일이익, 태양광, 유상증자, 하나대투증권, 수익성
Topic 14	에너지, 태양광, 태양전지, 시스템, 풍력발전, 태양광발전, 연료전지, 기술력, 에너지기술, 차세대, 폴리실리콘, 하이브리드, 도시가스, 실리콘, 핵심기술, 국산화, 카자흐스탄, 글로벌, 상용화, 연구소
Topic 15	에너지, 자동차, 성장동력, 차세대, 신소재, 바이오, 서비스, 시스템, 글로벌, 정보기술, 전략산업, 반도체, 제조업, 에너지산업, 중소기업, 발광다이오드, 의료기기, 미래형, 디스플레이, 일자리
Topic 16	간담회, 위원장, 공정위, 세미나, 공정거래, 감독원, 토론회, 시상식, 금감원, 공공기관, 국무총리, 중소기업, 활성화, 농림부, 공모전, 브리핑, 설명회, 코엑스, 수산식품, 개장안
Topic 17	에너지, 태양광, 연료전지, 풍력발전, 태양광발전, 바이오, 저탄소, 보급률, 태양열, 친환경, 에너지원, 생산량, 조성사업, 강원도, 청정에너지, 시범단지, toe, 활성화, 부산시, 폐기물
Topic 18	에너지, 태양광, 전시회, 그린에너지, 엑스포, 중소기업, 세미나, 대기업, 국내외, 에너지산업, 현대중공업, 글로벌, 아시아, 에너지관리, 상담회, 전문가, 박람회, 해외진출, 비즈니스, 자동차
Topic 19	에너지, 태양광, 에너지산업, 연료전지, 테스트, 전문가, 연구소, 활성화, 특성화, 인력양성, 인재양성, 전문인력, 프로그램, 전북대, 경남대, 캠퍼스, 중소기업, 성장동력, 클러스터, 에너지관리
Topic 20	에너지, 무할당제, 태양광, 사업자, 발전량, 조력발전, 인증서, 에너지공급, 발전사, RPS, 태양광발전, 발전설비, 공기업, 활성화, 대규모, 한국전력, 에너지원, 일정비율, 자회사, 지원제도

진출사업, 스마트 그리드 사업, 정부 주도 포럼, 온실가 스감축 방안이 핵심적인 분야라고 파악하였다.

또한, Table 1의 각 토픽들에 대한 연구 주제를 최근 10년간의 연도별 추이를 파악해보았다. 지속적인 관심으로 상승세를 보이는 5가지의 토픽주제를 확인하였다.

첫째, 스마트 그리드 사업으로 현재 공공기관과 기업들이 협력하여 전력과 IoT, 빅 데이터를 융합한 스마트 그리드 사업을 추진하고, 향후 에너지 신산업 활성화의 기회로 기대하고 있다. 또한 에너지 신산업은 스마트 그리드 구축 없이는 장기적으로 지속할 수 없기 때문에 스마트 그리드에 대한 관심과 지원과 함께 상승세를 보이고 있음을 확인할 수 있다(Topic 5).

둘째, 태양광 보급 사업이다. 현재 수준에서는 기존 화석에너지를 이용한 발전에 비하여 경제성이 상대적으로 낮으므로 환경문제와 미래 에너지의 안정적 확대를 배경으로 환경문제 해결, 기후변화 협약 대응에 의해 정부 주도형 산업이고, 태양광 보조금 정책 강화로 공공건물 옥상에 발전설비 사업을 지속적으로 확대되고 있다 (Topic 8).

셋째, 신재생에너지 공급 의무화 제도(RPS)는 국내 전기사업자들의 전력공급의 일정 부분을 의무적으로 신재생에너지로부터 생산해야 하는 제도이다. 공급 의무화 제도는 기존의 설비 중심이었던 태양광이나 풍력 발전 외에 다른 분야의 신재생에너지들의 사용 가능성이 크

며, 발전 사업자는 적은 비용으로 신재생에너지 발전을 위해 기존에 소외되었던 다양한 신재생에너지 시장이 형성되고 있으며 기존의 투자된 발전설비를 사용 가능성이 증가되고 있다(Topic 20).

넷째, 에너지 절약과 효율적인 에너지 사용을 위한 전력부문의 에너지 절감 교육프로그램으로 신재생에너지, 고효율에너지, 환경/온실가스를 주제로 다양한 프로그램들을 통해 다양한 분야의 사람들 또는 에너지에 관심을 가지고 있는 사람들에게 제공하고 있다(Topic 2).

다섯째, 자유무역협정(FTA)과 관련해 우리나라는 인프라, 보건의료, 신재생에너지 등의 고부가가치 신산업 분야를 중심으로 협력의 범위를 확대하고 있다. 최근 페루, 온두라스 등의 나라들과 협력을 통해서 경제성장파 에너지 관련 사업에 대한 활성화에 노력하고 있다(Topic 10).

최근 10년 동안의 증가세를 보이는 주제어는 녹색기술, 신재생에너지, 친환경 생활이다. 생활 곳곳에서 에너지 효율성과 에너지 절감을 위한 노력이 이루어지고 있어 신재생에너지에 대한 국민들의 관심이 증가하고 있는 것으로 판단됨을 토픽모델링 분석 결과를 통해 확인할 수 있다[Fig 5].

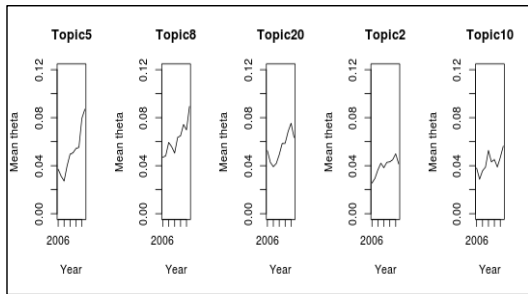


Fig. 5. An Topic modeling Analysis result of Hot Topic

신재생에너지에 대한 하강세를 보이는 분야를 확인하였다.

첫째, 선진국의 온실가스 감축을 위한 현황이다. 산업 발전에 따라 기후변화를 일으키는 유해가스의 배출이 증가될 수밖에 없다. 국가 간의 이해관계가 다르기에 소규모의 국가의 의지만으로 가능한 것은 아니다. 현재 기후 변화협상은 선진국과 개발도상국의 입장 차이로 지지부진하다. 또한 미국이 온실가스 배출량이 국가가 참여하지 않고는 온실가스 감축문제를 해결하기 어려울 것이다(Topic 4).

둘째, 신재생에너지 시장 활성화는 2020년까지 중장기적인 전망을 가지고 계획을 수립하고 있고 세계 신재생에너지 시장이 불황에서 벗어난 것과 달리 국내 기업의 체감은 여전히 낮은 상황이다(Topic 13).

셋째, 지역 경제 활성화(새만금 신재생에너지 단지)는 국가의 기후 변화협약에 대응한 온실가스 배출량 감축의무를 이행하기 위한 전략적 목표를 달성하고 신재생에너지 산업을 활성화 시키는 목적을 가지고 있다. 신재생에너지 융복합 지원 사업을 통해 구역 단위의 에너지 자립을 지원하는 동시에 보급 확대를 추진하고 문제점이 발생되고 있다. 신재생에너지 발전 사업에 있어 초기 투자 비용 문제, 민간사업자가 참여하기 어려운 문제, 발전설비 주민의 수용성 저하 등으로 산업 활성화에 대한 문제점이 있다. 신재생에너지에 대한 관심과 국가적 지원이 확대가 되어야 지역 경제 활성화에 대한 관심도가 증가될 것으로 보인다(Topic 11).

넷째, 청정개발체제(CDM:Clean Development Mechanism)은 자본과 기술을 가진 선진국이 개도국에 투자하고 사업을 진행하여 국제적 온실가스 저감 방안이다. 선진국과 개도국 관계에서 서로의 이익을 주고받을 수 있다. 개도국인 우리나라는 현재까지는 우리나라의 온실가스 저감에 대한 현실적인 의무를 받지 않고 있다. 신재생에너지에 대한 정부의 많은 관심을 넘어 현실적인 투자가 필요하다(Topic 3).

다섯째, 신재생에너지는 화석에너지 고갈에 따라 에너지 가격 폭등에 대한 대책으로써 주목받고 있는 에너지인 만큼 유가가 하락하면 신재생에너지 수요가 감소할 것이라는 우려가 폴리실리콘 가격 하락으로 이어지고 있다. 유가 하락이 지속하면 태양광 산업계 전체가 어려울 것이다(Topic 14).

최근 10년 동안의 하강세를 보이는 주제어는 기후변화, 에너지 투자이다. 정부에서는 안정적인 전력 공급과 원자력발전소 조성 정책을 추진하였으나, 후쿠시마 사고, 원전발전소 사고로 인해 원자력발전에 대한 신뢰도의 하락과 지역주민들의 반대가 큰 요인으로 작용한 것으로 판단된다[Fig.6].

Table 1에서 제시된 신재생에너지 관련 사회적 동향은 과거에는 신재생에너지 온실감축을 위한 다양한 사업이었다면, 최근에는 ICT를 접목시켜서 새로운 에너지원 보다는 정책적, 이미 개발된 기술을 활용한 움직임이 보인다.

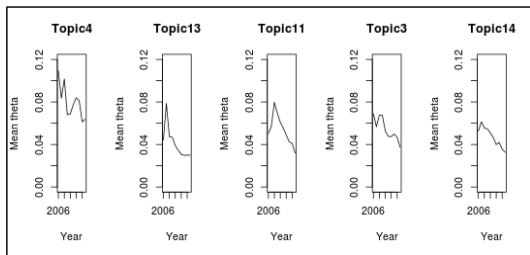


Fig. 6. An Topic modeling Analysis result of Cold Topic

### 5. 결론

본 연구에서는 빅데이터 분석 기법들로 신재생에너지 분야에 대한 연구동향을 파악하기 위해 한글문서인 언론 기사를 텍스트 마이닝 기법과 토픽모델링 기법의 분석방법을 통해 분석하였다. 연구 결과를 바탕으로 신재생에너지 분야의 전반적인 동향을 정리해보았다. 첫째 신재생에너지에 대한 필요성과 중요성이 대두되면서 새로운 에너지 확보를 위한 정책적으로 노력이 나타나고 있으며, 신재생에너지 산업화를 위해 정책적 지원이 활발히 진행 중임을 알 수 있었다. 둘째, 신재생에너지 관련 언론 기사를 대상으로 20개의 주제를 발견 할 수 있었다. 우선 신재생에너지 시장에 대한 관심으로 인한 국가와 기업이 함께 신재생에너지 사업인 융복합분야 사업화, 신재생에너지 산업 활성화를 위한 사업에 대한 토픽 주제가 가장 많이 나타났다. 정보통신(ICT)와 에너지 사업을 접목하는 스마트 그리드 전력망 사업에 대한 활용이 중요한 주제로 부상되고 있음을 확인할 수 있었다. 이러한 연구동향을 바탕으로 향후 신재생에너지 연구는 ICT 기술을 바탕으로 신재생에너지에 대한 기술적인 측면이 활발히 적용될 것으로 예측된다.

이 연구는 다양한 분야에서 부상 중인 신재생에너지의 언론기사만 수집하여 연구가 진행되었기에 에너지에 대한 전반적인 동향만 파악할 수 있었다. 신재생에너지에 대한 연구는 지속적으로 관심을 가지고 이루어지고 있음에 연구 동향 분석을 제공하여 향후 추가적인 연구에 기초자료로 활용되고 에너지 분야의 발전에 기여할 것으로 기대된다. 또한 다양한 에너지원을 통해서 신재생에너지에 대한 연구 동향을 파악한다면 보다 정교한 유망 주제 도출이 가능할 것으로 기대된다.

### References

- [1] Blei, D. M., Ng, A. Y., & Jordan, M. I. "Latent dirichlet allocation". the Journal of machine Learning research, 3, 993-1022. 2003.
- [2] Blei, D. M., & Lafferty, J. D. Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). ACM. 2006.  
DOI: <http://dx.doi.org/10.1145/1143844.1143859>
- [3] Seol A Jin, Go Eun Heo, Yoo Kyung Jeong, & Min Song. "Topic-Network based Topic Shift Detection on Twitter". Korea Society for Information Management, 30(1), 285-302. 2013.  
DOI: <http://dx.doi.org/10.3743/KOSIM.2013.30.1.285>
- [4] Ja-Hyun Park, & Min Song. "A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling". Korea Society for Information Management, 30(1), 7-32. 2013.  
DOI: <http://dx.doi.org/10.3743/KOSIM.2013.30.1.007>
- [5] Griffiths, T. L., & Steyvers, M. Finding scientific topics. Proceedings of the National Academy of Sciences, 101(suppl 1), 5228-5235.2004.  
DOI: <http://dx.doi.org/10.1073/pnas.0307752101>
- [6] Mimno, D. M. and McCallum, A. "Topic models conditioned on arbitrary features with Dirichlet-multinomial regression," 411~418. 2008
- [7] Hornik, K., & Grün, B. topicmodels: An R package for fitting topic models. Journal of Statistical Software, 40(13), 1-30. 2011.
- [8] Steyvers, M., & Griffiths, T. Probabilistic topic models. Handbook of latent semantic analysis, 427(7), 424-440. 2007.
- [9] Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.  
DOI: <http://dx.doi.org/10.1145/2133806.2133826>
- [10] Shi, C., Cai, Y., Fu, D., Dong, Y., & Wu, B. "A link clustering based overlapping community detection algorithm". Data & Knowledge Engineering, 87, 394-404. 2013.  
DOI: <http://dx.doi.org/10.1016/j.datak.2013.05.004>
- [11] Pons, P., & Latapy, M. Computing communities in large networks using random walks. In Computer and Information Sciences-ISCIS 2005 (pp. 284-293). Springer Berlin Heidelberg. 2005.  
DOI: [http://dx.doi.org/10.1007/11569596\\_31](http://dx.doi.org/10.1007/11569596_31)
- [12] Newman, M. E., & Girvan, M. Finding and evaluating community structure in networks. Physical review E, 69(2), 026113. 2004.  
DOI: <http://dx.doi.org/10.1103/PhysRevE.69.026113>

**신 규 식(Kyu-Sik Shin)**

[준회원]



- 2014년 3월 ~ 현재 : 고려대학교  
산업경영공학부 석사과정

<관심분야>

빅데이터, 데이터마이닝, 자연어처리

---

**최 회 련(Hoe-ryeon Choi)**

[정회원]



- 2003년 3월 ~ 현재 : 고려대학교  
산업경영공학부 박사과정

<관심분야>

생산 및 정보 관리 시스템, AI, 온톨로지

---

**이 흥 철(Hong-Chul Lee)**

[정회원]



- 1983년 2월 : 고려대학교 산업공학  
학사
- 1988년 2월 : Univ. of Texas 산업  
공학과 석사
- 1993년 2월 : Texas A&M Univ.  
산업공학박사
- 1996년 3월 ~ 현재 : 고려대학교  
산업경영공학부 교수

<관심분야>

SCM, 생산 및 물류 정보시스템, PLM