

텍스트마이닝 기법을 활용한 사용후핵연료 건식처리기술 관련 언론 동향 분석

정지송

과학기술연합대학원대학교 양자에너지화학공학과
(jjsong9410@gmail.com)

김호동

한국원자력연구원 책임연구원
(khd@kaeri.re.kr)

최근 4차 산업혁명, 코로나로 인한 뉴노멀 시대의 도래 등을 계기로 인공지능, 빅데이터 연구와 같은 언택트 관련 기술의 중요성이 더욱 급상하고 있다. 각종 연구 분야에서는 이러한 연구 트렌드를 따라가기 위한 융합적 연구가 본격적으로 시행되고 있으나 원자력 분야의 경우 자연어 처리, 텍스트마이닝 분석 등 인공지능 및 빅데이터 관련 기술을 적용한 연구가 많이 수행되지 않았다. 이에 원자력 연구 분야에 데이터 사이언스 분석기술의 적용 가능성을 확인해보고자 본 연구를 수행하였다. 원자로 연료로 사용된 뒤 배출되는 사용후핵연료 인식 동향 파악에 대한 연구는 원자력 산업 정책에 대한 방향을 결정하고 산업정책 변화를 사전에 대응할 수 있다는 측면에서 매우 중요하다. 사용후핵연료 처리기술은 크게 습식재처리 방식과 건식재처리 방식으로 나뉘는데, 이 중 환경 친화적이고 핵비확산성 및 경제성이 높은 건식재처리 기술인 '파이로프로세싱'과 그 연계 원자로 '소듐냉각고속로'의 연구개발에 대한 재평가가 현재 지속적으로 검토되고 있다. 따라서 위와 같은 이유로, 본 연구에서는 사용후핵연료 처리기술인 파이로프로세싱에 대한 언론 동향 분석을 진행하였다. 사용후핵연료 처리기술인 '파이로프로세싱' 키워드를 포함하는 네이버 웹 뉴스 기사 전문의 텍스트데이터를 수집하여 기간에 따라 인식변화를 분석하였다. 2016년 발생한 경주 지진, 2017년 새 정부의 에너지 전환정책 시행된 2010년대 중반 시기를 기준으로 전, 후의 동향 분석이 시행되었고, 빈도분석을 바탕으로 한 워드 클라우드 도출, TF-IDF(Term Frequency - Inverse Document Frequency) 도출, 연결정도 중심성 산출 등의 분석방법을 통해 텍스트데이터에 대한 세부적이고 다층적인 분석을 수행하였다. 연구 결과, 2010년대 이전에는 사용후핵연료 처리기술에 대한 사회 언론의 인식이 외교적이고 긍정적이었음을 알 수 있었다. 그러나 시간이 흐름에 따라 '안전(safety)', '재검토(reexamination)', '대책(countermeasure)', '처분(disposal)', '해체(disassemble)' 등의 키워드 출현빈도가 급증하며 사용후핵연료 처리기술 연구에 대한 지속 여부가 사회적으로 진지하게 고려되고 있음을 알 수 있었다. 정치 외교적 기술로 인식되던 사용후핵연료 처리기술이 국내 정책의 변화로 연구 지속 가능성이 모호해짐에 따라 언론 인식도 점차 변화했다는 것을 확인하였다. 이러한 연구 결과를 통해 원자력 분야에서의 사회과학 연구의 지속은 필수불가결함을 알 수 있었고 이에 대한 중요성이 부각되었다. 또한, 현 정부의 원전 감축과 같은 에너지 정책의 영향으로, 사용후핵연료 처리기술 연구개발에 대한 재평가가 시행되는 이 시점에서 해당 분야의 주요 키워드 분석은 향후 연구 방향 설정에 기여할 수 있을 것이라는 측면에서 실무적 의미를 갖는다. 더 나아가 원자력 공학 분야에 사회과학 분야를 폭넓게 적용할 필요가 있으며, 국가 정책적 변화를 고려해야 원자력 산업이 지속 가능할 것으로 사료된다.

주제어 : 사용후핵연료, 텍스트마이닝, 워드 클라우드, TF-IDF, 의미 네트워크 분석

논문접수일 : 2021년 2월 8일 논문수정일 : 2021년 6월 1일 게재확정일 : 2021년 6월 24일

원고유형 : 일반논문 교신저자 : 김호동

1. 서론

2007년 스마트폰 혁명을 계기로 빅데이터 시대가 본격적으로 도래하였으며 잇다른 IT 혁신으로 디지털 휴대용 기기의 사용량이 급증하게 되었다. 디지털 휴대용 기기의 대중화로 인해 소셜 네트워크 서비스(SNS), 메신저 서비스, 클라우드 컴퓨팅 등 새로운 IT 비즈니스가 창출되었고 실시간 비정형데이터의 창출 속도 또한 급상승하였다. 우리 주변에 존재하는 모든 형태의 데이터 중, 약 80% 이상이 비정형데이터일 정도로 비정형데이터에 대한 그 중요성이 높아지고 있다 (Chakraborty and Pagolu, 2014).

비정형데이터 분석 방법 중 가장 널리 사용되는 분석법은 텍스트마이닝 네트워크 분석이다. 텍스트마이닝이란, 언어학, 통계학, 기계학습 등을 기반으로 한 자연어 처리기술을 활용하여 텍스트데이터를 정형화하고, 특징을 추출하기 위한 기술과 추출된 특징으로부터 의미 있는 정보를 발견할 수 있도록 하는 기술이다. 텍스트 네트워크 분석은 네트워크의 시각화를 통한 현상과 구조해석으로 대상에 대한 지식과 이해를 증진시키고, 의미 형성에 중요한 영향력이 되는 분석 결과를 통해 정량적·통계적으로 그 의미와 흐름을 파악할 수 있다.

대규모의 방대한 데이터가 실시간으로 생성되는 4차 산업혁명 시대의 흐름에 따라 의료, 제조, 금융, 교육, 과학기술 등의 다양한 분야에서 인공지능 관련 기술이 연구되고 있으며 이를 활용한 서비스 수가 비약적으로 증가하고 있다. 그러나 원자력 분야에서는 빅데이터 연구 방법을 적용한 연구사례가 많지 않다. 특히나 사회적으로 최근 몇 년 동안 원자력 관련 이슈가 끊임없이 논란화 됨에도 불구하고, 원자력 관련 언론 동향이 어떤지에 대한 분석연구는 아직 미비한 편이다.

이에 본 연구는 통계적이고 정량적 성격을 띠는 텍스트마이닝 분석기술을 적용하여, 원자력 관련 사회적으로 가장 중점적으로 다뤄지고 있는 ‘사용후핵연료 건식처리기술’에 대한 언론의 인식 변화를 객관적으로 분석해보고자 하였다. 사용후핵연료 건식처리기술의 언론 동향 분석은 국내 대표 포털 사이트인 네이버에서 파이썬 코드를 통해 수집한 뉴스 기사를 바탕으로 워드 클라우드, TF-IDF (Term Frequency - Inverse Document Frequency), 연결정도 중심성 값 산출 및 도식화를 통해 주요 키워드 분석 및 키워드 간 상관관계를 분석하였다.

2. 관련 연구

2.1 빅데이터 분석기술

빅데이터 분석 방법은 분석하고자 하는 데이터의 형태에 따라 그 분석 방법이 나누어진다. 고정된 필드에 수치로 저장된 데이터인 정형 데이터의 경우, 명목척도, 서열척도 등의 개념을 적용해 회귀분석과 같은 통계적 방법을 통해 분석이 수행된다. 반면, 고정된 필드에 저장되어 있지 않은 동영상, 이미지, 텍스트데이터와 같은 비정형데이터의 경우 또다시 세부적으로 분석 방법이 나뉜다.

영상이나 이미지 데이터의 경우 영상인식을 위한 지도 학습기술, 비지도 학습기술, 강화학습 등의 분석기술이 사용되게 된다 (J.G.Ko, 2014; S.Y.Kim, Y.J.Chung, 2017; Y.S.Park, 2019). 음성 및 텍스트 데이터의 경우는 확보된 텍스트를 기반으로 분석이 수행되게 된다. 텍스트는 가장 기본적이고 광범위한 비중을 차지하는 비정형데이터이고, 이러한 텍스트를 컴퓨터 공학, 통계학, 사회학 등의 학문적 지식을 이용하여 특정 목적에 맞게 유의미한 정보

를 추출하는 분석 및 처리 과정을 텍스트마이닝 (Textmining) 이라고 한다.

텍스트마이닝은 텍스트가 내포한 핵심 의미 추출 방법을 정형화시키기 위해 수십 년간 다학제적 연구가 이루어진, 비정형데이터 중 가장 핵심적이고 필수적인 기술이라고 할 수 있다. (J.Y.Jang, 2013) 현재 통용되고 있는 텍스트마이닝 분석 방법으로는 형태소 분석, 벡터 공간 모델, 의미 네트워크 분석, 동시 출현단어, 텍스트 감성분석 등이 존재한다 (Statistics Training Institute, 2017). 형태소 분석의 경우 텍스트를 분석하는 가장 보편적인 방법으로, 의미의 최소 단위인 형태소를 통해 품사 태깅, 단어 식별 등을 할 수 있다. 벡터 공간 모델 분석이란 머신러닝에서 다수 사용되며 문서(document)와 단어 (tem)를 각각 벡터(vector) 차원에 대응시켜 통계적 가중치를 구하는 개념이다. 이 분석기술은 TF-IDF (Term Frequency - Inversed Document Frequency), 카이제곱 검정, 코사인 유사도 등을 통해 단어의 빈도수 분포에서 중요하지 않은 단어를 걸러내고, 문서의 유사도를 구할 수 있다는 장점이 있다. (Salton G, 1975) 의미 네트워크 분석이란 키워드 간의 의미 혹은 맥락상의 연결 관계를 정의하고, 해당 연결 관계를 시각화하거나 중요한 개념들을 네트워크 속의 중심성 정도(centrality)에 따라 추출하는 방법이다. 그 외에도 단어 사이 의미 관계성을 파악하기 위한 동시 출현단어 분석, 감성 사전을 기반으로 텍스트의 감성 비율을 정량화하는 텍스트 감성분석 등의 분석 기술이 존재한다 (Staticstis Training Institute, 2017).

2.2 원자력 연구분야에서의 빅데이터 분석기술 적용 사례

사회과학 분야에서 텍스트데이터 분석을 통한 텍

스트마이닝 연구사례는 다수 존재한다. 텍스트상에 출현한 키워드 간의 연결과 연관성 정도를 분석할 수 있는 키워드 네트워크 분석은 초기에 문헌정보 학계를 중심으로 연구되었으나, 최근에는 언론학, 행정학, 경영학, 공학 등 다양한 학문에서 네트워크 구조를 분석하기 위한 유용한 방법론으로 널리 받아들여지고 있다 (D.R.Kim, 2015).

원자력 연구 분야에서 텍스트데이터 분석방법이 적용된 것은 그리 오래되지 않았다. 김동성(2015)은 소셜 미디어 트위터 (twitter) 상에서 2009년부터 2013년까지의 원자력 관련 비정형 텍스트데이터를 바탕으로 원자력 관련 여론의 긍정/부정 판별을 위한 감성분석을 수행하였다. 안재훈(2020)은 신고리 5, 6호기 공론화 과정에서의 내용 및 대립 요인을 분석하기 위해 공론조사 시 전문가가 ‘시민참여 대표단’에 제공한 발표문과 이에 대한 질의응답 총 16 개의 도큐먼트를 대상으로 빈도분석, 연결정도 중심성 분석 등을 수행하였다.

그러나 앞서 언급한 연구는 객관적인 데이터가 아닌 주관적 성격이 뚜렷한 트위터 또는 질의응답 내용 텍스트데이터를 연구 대상으로 삼았으며,

연구분야에 대한 적용이 아닌 원자력을 주제로 한 발표문을 분석한 연구라는 한계가 있다. 또한, 한국어의 경우 인터넷 비속어, 약어, 은어, 비꼬는 말 등이 다수 존재하기에, 비정형 텍스트 데이터 분석 방법으로 텍스트 감성분석을 적용시키는 것은 정확도 높은 결과 도출이 어렵다.

위와 같은 선행연구 결과, 본 연구는 비정형데이터를 분석/활용할 수 있는 기술 요구가 점차 증대되고 있는 현대 흐름에 맞춰 빅데이터 분석방법인 텍스트마이닝 을 접목해 원자력 분야에서 수행되었던 기존연구들의 한계점을 보완하고자 한다. 이에 사용후핵연료 건식처리기술 관련 비정형 데이터를 바탕으로 분석을 진행하였고 언론의 인식 동향에

대해 살펴보고자 하였다.

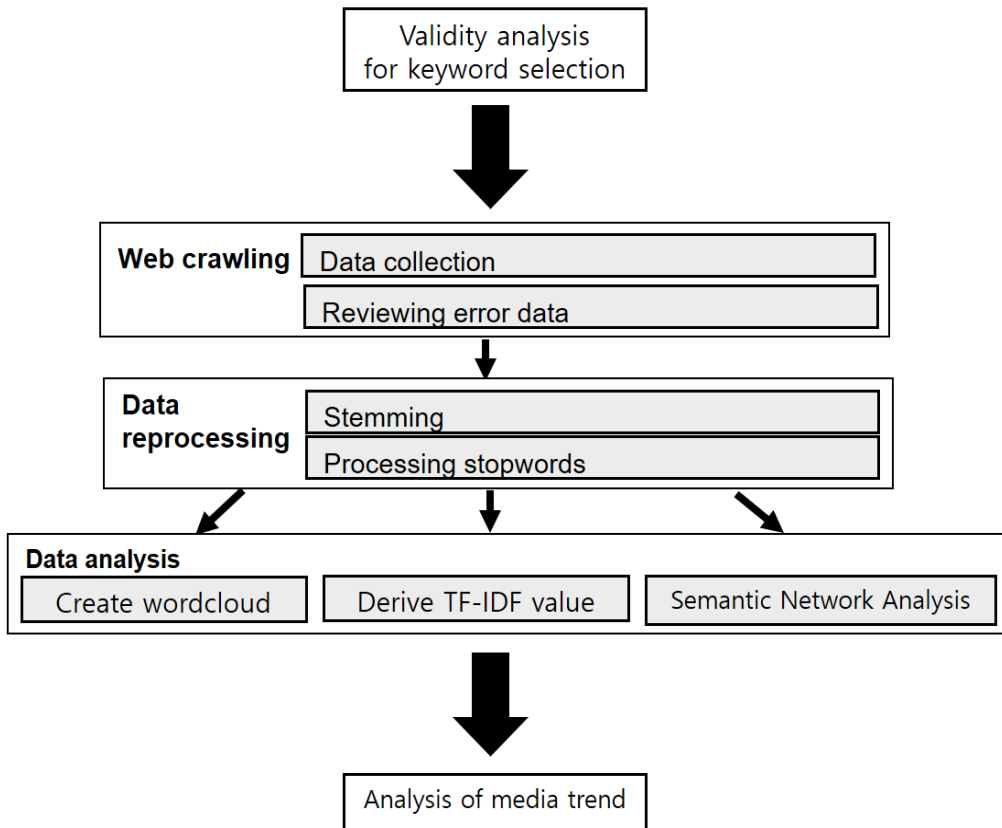
국가 기술 데이터인 사용후핵연료 처리기술 관련 연구 분야 특성상, 정형 데이터, 영상 및 이미지 데이터 확보 어려움 등의 상황을 고려하여 수집이 비교적 용이한 포털 사이트의 텍스트 데이터를 연구 대상으로 선정하였다. 이를 바탕으로 텍스트마이닝은 형태소 분석과 벡터 공간 모델, 동시 출현단어 분석 방법을 적용해 단어 빈도분석, 워드 클라우드 작성 및 TF-IDF 분석을 수행하였다. 이 외에도 의미 네트워크 분석을 추가로 수행하여 더욱 심도 있고 다차원적인 텍스트마이닝 분석을 시행하였다. 이는 원자력 관련 언론의 인식 변화 분석을 위해 텍

스트 기반의 빅데이터 분석법을 적용하여 정량적 분석을 수행한 새로운 접근 사례를 가진다는 의의를 떨 것으로 판단된다.

3. 데이터 수집 및 분석방법

3.1 웹 텍스트데이터 수집

본 연구는 <Figure 1>과 같은 순서로 진행되었다. 사용후핵연료 건식처리기술에 대한 언론 동향 분석을 위해 국내 대표 온라인 포털 사이트인 네이



<Figure 1> Research Procedure

비)에서 파이썬 코드를 통해 사용후핵연료 처리 기술인 ‘파이로프로세싱’과 그 연계 원자로인 ‘소듐냉각고속로’ 관련된 기사를 수집하였다. 수집된 데이터는 키워드가 기사 내 처음으로 출현하였던 2005년 12월 20일부터 2020년 3월 25일까지의 뉴스 정보였으며, 해당 기사의 제목, 출간지, 출간일자, 전문, url 등의 텍스트데이터를 모두 크롤링하였다. 그 결과, ‘파이로프로세싱’ 키워드를 포함하는 기사는 총 1,660건이었고, ‘소듐냉각고속로’ 키워드 포함 기사는 754건이었다. 이 중 중복되는 데이터는 총 436건이었으며 이를 정리한 결과, 최종적으로 총 1,978건의 데이터를 대상으로 분석을 진행하였다. 1,978건의 데이터는 인문사회분야의 연구 측면에서 적은 양의 데이터일 수 있지만, 기술적인 성격이 강한 원자력 관련 분야에서는 유의미한 수의 데이터라고 할 수 있다. 이렇게 수집된 데이터를 바탕으로 형태소 분석(Stemming)²⁾, 불용어(Stopwords)³⁾ 처리 등의 데이터 전처리과정을 거쳐 빈도분석, TF-IDF 분석, 의미 네트워크 분석을 시행하였다.

3.2 웹 텍스트데이터 분석방법

비정형데이터를 처리하는 빅데이터 분석방법 중 가장 대표적인 텍스트마이닝 기법으로는 워드 클라우드

분석, TF-IDF, 의미 네트워크 분석방법 등이 있다. 본 연구에서는 다각도적인 분석을 위해 빈도 분석, TF-IDF 분석, 의미 네트워크 분석을 차례로 수행하였다.

3.2.1 워드 클라우드

워드 클라우드는 텍스트데이터를 구성하고 있는 키워드들의 출현 빈도수를 기반으로 핵심 단어 크기에 비례하여 시각화하는 분석방법이다(Lim, DH, 2015). 이는 가시성이 높은 요약물을 제공하기 때문에 텍스트데이터의 개념을 직관적으로 파악할 수 있다는 특징을 갖는다. 워드 클라우드는 파이썬(Python)의 라이브러리인 wordcloud⁴⁾와 matplotlib⁵⁾ 을 통해 작성하였다.

3.2.2 TF-IDF(Term Frequency - Inverse Document Frequency)

TF-IDF(Term Frequency - Inverse Document Frequency)란 어떤 단어가 문서 내에서 얼마나 중요한 위치를 차지하는지 그 정도를 나타내는 통계적 지표이다. 어떤 문서에 특정 단어가 자주 출현한다면, 해당 단어는 그 문서와 연관성이 높다고 말할 수 있을 것이나, 무조건적으로 출현 빈도에 의존하

- 1) 웹 분석 통계자료인 <인터넷 트렌드> 에 따르면 2020년 12월 국내 포털사이트 점유율 결과, 네이버(Naver)의 국내 포털 점유율은 58.31%로 1위를 차지하였고 그 뒤로 구글(Google)이 34.24%로 2위를 차지했다. 즉, 많은 사용자가 정보 검색 목적으로 ‘네이버’ 사이트를 가장 많이 사용하는 것을 알 수 있었고 국내 언론 동향 분석을 위한 데이터 수집 사이트로 네이버가 가장 타당하다고 판단하였다.
- 2) 형태소 분석(Stemming)이란 어형이 변형된 단어로부터 접사 등을 제거하고 그 단어의 어간을 분리해내는 것이다. 예를 들어 진행형, 분사, 현재 완료형인 단어를 동사 원형으로 추출하여 내포하고 있는 뜻을 분석하기 위한 과정이다.
- 3) 불용어(Stopwords)는 텍스트 내에 자주 등장하지만 분석하는 것에 있어서는 큰 도움이 되지 않는 의미 없는 단어를 뜻한다. 조사, 접미사 등은 분석 시에 큰 의미를 띠고 있는 키워드가 아니기에 불용어 처리하여 데이터 분석에서 제외한다.
- 4) 워드 클라우드를 생성하기 위해 불러와야 할 파이썬 모듈로, 단어의 중요도에 따라 단어의 크기나 색을 변화시켜 어떤 단어가 중요한지 보여주는 시각화 라이브러리이다. 본 연구에서는 형태소 분석 및 불용어 처리가 완료된 데이터를 바탕으로 워드 클라우드를 작성하였다.
- 5) matplotlib이란 파이썬에서 가장 많이 쓰이는 대표적인 데이터 시각화 라이브러리로, 데이터를 차트나 플롯, 히스토그램 등의 다양한 스타일로 출력해낼 수 있다.

면 전체적인 분석의 정확도가 떨어진다. 따라서 분석과정 중, 어느 문서에서나 빈번하게 등장하는 단어는 연관성 측정에서 제외해야 하는데 이를 보완하기 위한 분석방법이 TF-IDF이다 (Y.H.Kim, 2008).

TF-IDF는 TF(Term Frequency)와 IDF (Inverse Document Frequency) 값을 곱한 값으로 TF는 단어 빈도수를 의미하여 이는 특정 단어가 한 문서 내에 얼마나 자주 등장하는지 빈도수를 나타내는 값이다. IDF (Inverse Document Frequency)란 역문서 빈도를 뜻하며, 문서빈도 DF (Document Frequency)에 역수 변환을 적용한 값이다. DF는 특정 단어가 몇 개의 문서에서 얼마나 자주 나타나는지를 측정하는 방법으로 특정 단어가 나타나는 문서의 수를 의미하며 단어가 등장하는 문서의 수가 클수록 그 단어가 사람들 사이에서 보편적으로 사용된다고 본다. 이 DF 값에 역수 변환을 해줌으로써 여러 문서에서 많이 등장하는 단어들에 대해 일종의 페널티를 주는 개념이 바로 IDF, 역문서 빈도이다 (<Formula 1> 참조).

$$u_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

$tf_{i,j}$ = number of occurrences of i in j

df_i = number of documents containing i

N = total number of documents

3.2.3 의미 네트워크 분석

의미 네트워크 분석(Semantic Network Analysis, SNA)은 키워드 간의 속성 및 관계 대표성을 파악하는 분석법이다 (Van Atteveldt W.H, 2008). 키워드들이 어떠한 단어들과 연관되어 있는지, 텍스트 데이터 내에 핵심 키워드는 무엇인지를 형성된 관

계망을 통해 파악할 수 있다. 네트워크에 대한 분석 방식은 크게 매개 중심성, 인접 중심성 그리고 연결 정도 중심성, 3가지로 나뉜다 (K,H,Eom, 2018).

가장 먼저, 매개 중심성이란 한 노드가 연결망 내의 다른 노드들 사이에 위치하는 정도를 측정하는 분석방법이다. 특정 노드가 주변 노드와의 네트워크를 구축하는 데 있어 중개자 혹은 다리 역할을 얼마나 수행하느냐를 측정하는 개념으로 키워드의 중개역할을 중심으로 분석을 시행할 때 고려되는 방법이다. 노드 사이의 최단 경로 위에 있는 노드일수록 매개중심성이 높다고 볼 수 있다 (Y.H.Kim, 2016).

두 번째로, 인접 중심성이란 각 노드(키워드)가 얼마나 네트워크 중앙에 위치하는지를 측정하는 값이다 (J.H.Han, 2015). 인접중심성은 키워드(노드)의 영향력 정도를 분석할 때 주로 고려되는 방법이며, 인접 중심성이 높은 노드는 네트워크 전체에 정보를 빠르게 전달/확산하여 영향력이 가장 큰 노드라고 본다. 마지막으로, 연결정도 중심성은 한 노드에 연결된 다른 노드들의 개수(연결정도, degree)를 기준으로 연결의 중심을 파악할 수 있는 값이다. 다른 노드들과 많이 연결되어있는 노드일수록 그 중요도가 높다고 판단되며, 이러한 특징으로 연결정도 중심성은 특정 주제에 대한 동향 분석 시에 가장 유용하게 사용되는 분석값이다 (S.S.Lee, 2016). 연결정도 중심성을 산출하는 식은 <Formula 2>와 같다.

$$\frac{\text{degree}}{\text{Max. of degree}} = \frac{\text{degree}}{n-1} \quad (2)$$

degree = Number of nodes connected with

n = Number of nodes

Max. of degree = Maximum number of nodes can be connected

3.2.3.1. 내향연결 중심성 (In-degree Centrality)

연결정도 중심성은 노드의 링크 방향에 따라 내향연결 중심성과 외향연결 중심성의 2가지로 분류된다 (Y.H.Kim, 2016). 내향연결 중심성은 노드로 들어오는 방향의 연결들에 대해 해당 노드의 중심성 수치를 계산하는 값이다. 통상적으로 내향연결 중심성은 핵심 키워드(노드)의 인기 정도로 해석되며 도출식은 <Formula 3>와 같다.

$$degree_{ik} = \sum_{i=1}^N Z_{ijk} = Z_{jk} \quad (3)$$

: Number of degrees that I receives from all other nodes(j) in network K

3.2.3.2. 외향연결 중심성 (Out-degree Centrality)

외향연결 중심성은 내향연결 중심성과 반대되는 개념으로, 노드에서 밖으로 나가는 방향의 연결에 대해 중심성 수치를 계산하는 값이다. 외향연결 중심성은 각 노드들(키워드)의 활동성 수치로 해석되며 도출 식은 <Formula 4>와 같다.

$$out-degree_{ik} = \sum_{i=1}^N Z_{ijk} = Z_{jk} \quad (4)$$

: Number of degrees from I to all other nodes(j) in network K

이에 본 연구는 시간 흐름에 따른 사용후핵연료 건식처리기술 ‘파이로프로세싱’에 대한 언론의 인식 변화를 분석하기 위하여 내향연결 중심성

(In-degree Centrality)과 외향연결 중심성(Out-degree Centrality) 값을 도출하였다. 도출하기 위한 툴로는 넷마이너(NetMiner)⁶⁾ 프로그램을 사용했으며, 도출된 값을 바탕으로 기간에 따른 노드 간 연결정도 중심성 그래프를 작성하였다.

4. 분석결과

사용후핵연료 건식처리기술인 ‘파이로프로세싱’에 대한 사회 언론의 인식 변화를 파악하기 위해 ‘파이로프로세싱’ 또는 ‘소듨냉각고속로’ 키워드를 포함하고 있는 네이버 뉴스 기사의 텍스트데이터를 바탕으로 빈도분석, TF-IDF 분석, 의미 네트워크 분석을 시행하였으며 분석결과는 다음과 같다.

4.1. 빈도분석 결과

<Table 1>은 ‘파이로프로세싱’, ‘소듨냉각고속로’ 관련 첫 기사가 보도된 2007년부터 2020년까지의 연도별 출현빈도 상위 10개 한글 키워드 순위이다. 2007년부터 2015년까지는 ‘개발(development)’, ‘협정(agreement)’, ‘한미(ROK-US)’, ‘양국(two nations)’ 등의 긍정적인 키워드들이 상위에 나타났지만, 2016년부터 2020년까지는 ‘안전(safety)’, ‘검토(examination)’, ‘정부(government)’, ‘처분(disposal)’ 등의 키워드가 상위권에 도출되면서 사용후핵연료 처리방안에 대한 사회인식이 변화되었음을 확인할 수 있었다.

연도별 키워드 출현빈도 분석결과를 바탕으로 2007년부터 2010년대 중반, 2010년대 중반부터

6) 넷마이너(NetMiner)란 사이람에서 개발한 사회연결망 분석을 위한 대표적인 소프트웨어이다. 관계 데이터에 대한 다각도적 분석과 시각화 작업을 수행할 수 있으며 직관적인 인터페이스와 관계 데이터의 시각화, 분석, 변환, 통계, 차트 등의 기능을 통합적으로 제공한다. 대표적인 기능으로는 텍스트데이터에서 키워드를 추출하여 추출된 단어 간의 네트워크를 구성하여 연관 단어, 핵심 단어, 세부 주제를 파악할 수 있는 의미망 분석 기능이 있다.

<Table 1> Term Frequency by year

Rank	2007	2008	2009	2010	2011	2012	2013
1	development	development	development	agreement	development	development	agreement
2	roadmap	roadmap	reactor	negotiation	plan	management	negotiation
3	plan	plan	uranium	revision	business	agreement	ROK-US
4	government	government	government	problem	initiative	reactor	enrichment
5	reactor	reactor	agreement	development	reactor	cooperation	government
6	agreement	agreement	production	ROK-US	facility	disposal	problem
7	future	future	export	proliferation	sodium	facility	revision
8	ministry of science and ICT	ministry of science and ICT	revision	government	government	government	uranium
9	waste	waste	world	plutonium	cooled	radioactive waste	two nations
10	world	world	system	facility	agreement	uranium	cooperation
Rank	2014	2015	2016	2017	2018	2019	2020
1	development	agreement	daejeon	safety	examination	energy	facility
2	agreement	ROK-US	facility	resident	business	government	storage
3	cooperation	enrichment	safety	development	committee	policy	government
4	consultation	uranium	development	business	safety	reactor	disposal
5	storage	revision	plan	daejeon	ministry of science and ICT	development	development
6	change	two nations	government	investigation	government	director	radioactive
7	problem	cooperation	committee	government	initiative	problem	waste
8	negotiation	negotiation	resident	citizen	development	business	operation
9	public opinion	export	experiment	budget	expert	advancement	radioactivity
10	government	consent	management	facility	national Assembly	construction	conference

2020년까지 두 기간으로 나눠 각각의 분석을 수행하였다. 2007년부터 2010년대 중반까지 보도된 기사 총 1,203건의 텍스트데이터를 기반으로 워드 클라우드 그래프를 도출하였다 (<Figure 2> 참조). 워드 클라우드 내 ‘협정(agreement)’, ‘협상(negotiation)’,

‘한미(ROK-US)’, ‘개정(revision)’, ‘개발(development)’ 등의 키워드의 크기가 크게 도출된 것으로 보아 해당 시기에 위 키워드들의 출현 빈도가 높다는 것을 알 수 있다. 또한 피이로프로세싱, 소듐냉각고속로와 같은 사용후핵연료 처리기술에 대해 정치 외교



<Figure 2> Word Cloud of NAVER news in 2007 to the mid 2010's



<Figure 3> Word Cloud of NAVER news in the mid to late 2010's

적인 기술로 인식되고 있음을 알 수 있다.

<Figure 3>는 2010년대 중반부터 2020년 현재까지 게재된 기사 775건의 텍스트데이터를 기반으로 워드 클라우드 빈도분석을 수행한 결과이다. 이전에 수행한 2007년부터 2010년대 중반 시기의 분석 결과 <Figure 2>와 비교해보았을 때, 워드 클라우드 내 '안전(safety)', '검토(examination)', '정부(government)',

'시설(facility)', '처분(disposal)', '예산(budget)', '검증(verification)', '금지(prohibition)' 등의 키워드가 다수 출현하고 있는 것을 알 수 있다. 이러한 점으로 보아, 사용후핵연료 처리기술의 안전성에 대해 시민들이 관심을 갖고 있고 이에 따라 정부의 예산정책 등에 대한 이슈가 증가한 것을 유추해볼 수 있다.

원자력계 큰 이슈 발생 시기에 빈도분석 결과를 대입해보았을 때, 2016년 경주 지진⁷⁾, 2017년 새 정부 출범에 따른 에너지 전환정책⁸⁾이 영향을 미친 것으로 보인다.

위 결과를 종합해서 고려해보았을 때, 2010년대 중반 시기를 기준으로 전 기간에는 사용후핵연료 건식처리기술에 대한 언론의 인식이 정치·외교적이며 기술 교류의 목적으로 긍정적으로 비추어지고 있었으나, 2010년대 중후반 시기부터는 처리기술 안전성 측면에서의 사회적 관심이 증가하였고 동시에 기술 지속성 여부에 대하여 재고되고 있음을 시사한다.

4.2 TF-IDF(Term Frequency - Inverse Document Frequency) 분석결과

2007년부터 2010년대 중반, 2010년대 중반부터 2020년까지의 기간별 출현 빈도수 및 TF-IDF에 따른 키워드 순위와 사용후핵연료 건식처리기술 관련 키워드 변화는 <Table 2>와 같다. 단어빈도만을 고려하였을 때는 두 기간 모두 상위권에 '사용후핵연료(nuclear spent fuel)', '기술(technology)', '원전(nuclear power plant)' 등과 같이 유사한 단어들 이 출현하며 뚜렷한 차이를 보이지 않는다. 그러나 TF-IDF 가중치 값을 보았을 때 키워드별 순위가 크게 뒤바뀌며 해당 기간 동안 어떠한 키워드가 높은

중요도를 가지고 있었는지를 확인해볼 수 있다.

2007년부터 2010년대 중반까지의 단어빈도 분석 결과, '협정(agreement)', '미국(US)', '사용후핵연료(nuclear spent fuel)', '기술(technology)', '재처리(reprocessing)' 등 한미 간 원자력 협정과 같은 외교적 의미를 대략적으로 유추해볼 수 있다. 동일시기의 TF-IDF 순위까지 함께 고려해보았을 때, '원자로(reactor)', '플루토늄(plutonium)', '안전(safety)', '동의(consent)', '방안(plan)', '비확산(nonproliferation)' 등의 키워드가 상위 출현함으로써 사용후핵연료 건식처리기술 및 한미 원자력 협정에 대한 세부적인 내용까지 확인해볼 수 있다.

2010년대 중반부터 2020년 현재까지의 단어빈도 분석 결과, '기술(technology)', '안전(safety)', '사용후핵연료(nuclear spent fuel)', '개발(development)', '파이로프로세싱(pyroprocessing)', '원전(nuclear power plant)' 등의 단어가 상위에 출현함으로써 보아 사용후핵연료 기술 관련 안전 문제가 대두된 것을 유추해볼 수 있다. 동일 기간의 TF-IDF 순위를 고려하였을 때는 '시민(citizen)', '예산(budget)', '정책(policy)', '처분(disposal)' 등의 단어가 새롭게 상위권에 출현한 것을 볼 수 있다. 즉 단어빈도 분석 결과만을 놓고 추론해보았을 때는 기술적인 문제가 쟁점이 되고 있음을 함의하고 있으나 TF-IDF 가중치 값까지 함께 고려하였을 때는 이것이 국내 사회·정책적인 문제로 변질하였음을 유추할 수 있다.

7) 2016년 9월 12일 경상북도 경주시 지역에서 발생한 리히터 규모 5.8 지진이다. 해당 지진은 1978년 대한민국에서 관측된 지진 이래로 역대 가장 강력한 지진이었으며, 이를 계기로 많은 국민들이 원전의 지진 안전성에 대해 관심을 가지게 되었다. (M.K.Kim, 2017)

8) 한국은 2년마다 향후 15년간의 안정적인 전력 수급을 위한 '전력수급기본계획'을 수립하여 미래 전력수요에 대비한다. 2017년 새 정부의 출범과 함께 2017년 12월에 발표된 '제8차 전력수급기본계획'에 따르면 향후 에너지 정책은 2031년까지 원전과 화석연료 사용을 꾸준히 줄여나가고 이를 대체하는 에너지원으로 LNG와 신재생에너지의 비중을 높여 시행될 것임을 공고하였다.

<Table 2> Term Frequency and TF-IDF analysis result by period

Rank	2007 - mid 2010's			
	Term Frequency		TF-IDF	
1	agreement	6,446	reactor	97.8
2	US	2,103	plutonium	97.6
3	nuclear spent fuel	5,274	safety	97.5
4	technology	5,177	consent	97.3
5	reprocessing	4,481	plan	97.1
6	research	4,400	world	97.1
7	nuclear power plant	3,775	supply	97.0
8	development	3,445	nonproliferation	97.0
9	negotiation	2,782	materials	97.0
10	revision	2,501	assessment	96.9
11	government	2,240	production	96.9
12	uranium	2,238	agreement	96.8
13	enrichment	2,226	progress	96.8
14	two nations	2,135	sodium	96.7
15	possibility	1,939	core	96.7
16	utilization	1,913	development	96.6
17	problem	1,869	cooled	96.6
18	nuclear fuel	1,846	storage	96.6
19	cooperation	1,826	future	96.6
20	facility	1,731	fast reactor	96.5
Rank	the mid to late 2010's			
	Term Frequency		TF-IDF	
1	technology	3,785	citizen	63.4
2	safety	2,978	nuclear fuel	63.4
3	nuclear spent fuel	2,900	budget	63.4
4	development	2,575	disposal	63.2
5	pyroprocessing	2,507	policy	63.2
6	nuclear power plant	2,504	experiment	63.0
7	institute	1,903	fast reactor	62.9

Rank	the mid to late 2010's			
	Term Frequency		TF-IDF	
8	business	1,674	future	62.9
9	govemment	1,399	US	62.9
10	daejeon	1,350	management	62.8
11	nuclear	1,283	area	62.8
12	reprocessing	1,079	nuclear	62.6
13	citizen	1,038	durability	62.6
14	plan	1,025	treatment	62.4
15	management	990	national assembly	62.3
16	problem	972	people	62.3
17	reactor	940	verification	62.1
18	nuclear fuel	909	radioactive waste	62.1
19	budget	882	expert	62.1
20	policy	837	facility	61.8

4.3 의미 네트워크 분석결과

키워드 빈도분석은 어떤 주제가 출현하고 사라졌는지에 대한 변화를 확인하는 데는 효과적이지만 단어의 존재 유무만으로는 표현되지 않는 주제를 파악하는 것에는 한계가 있다. (E.S.Yoo, 2015) 시간 흐름에 따른 사용후핵연료 건식처리기술 관련 언론의 인식 변화를 구체적으로 파악하기 위하여 의미 네트워크 분석(Semantic Network Analysis, SNA)을 실시하였다. 분석과정은 내향 연결정도 중심성, 외향 연결정도 중심성 값을 분석한 후 의미 네트워크 그래프 도출 순으로 진행되었다.

4.3.1 2007년 - 2010년대 중반 의미 네트워크 분석

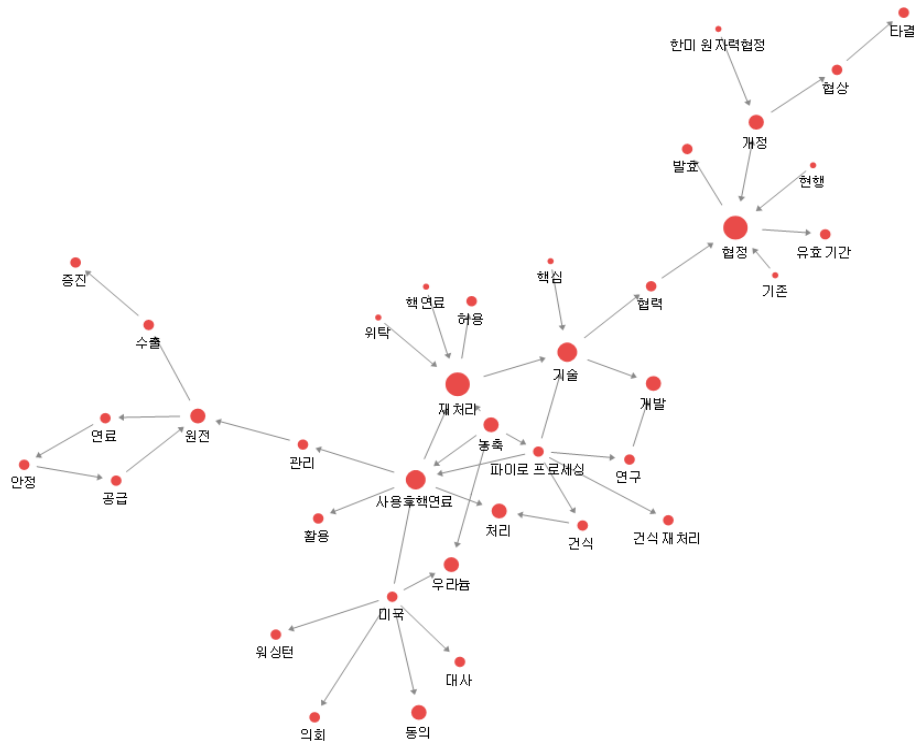
<Table 3>은 2007년부터 2010년대 중반까지 사용후핵연료 건식처리기술 관련 보도된 기사 총

1,203건의 텍스트데이터를 바탕으로 내향/외향 연결중심성을 분석한 결과이다. 내향 연결정도 중심성(In-Degree centrality) 수치 도출결과, ‘협정(agreement)’, ‘재처리(reprocessing)’, ‘사용후핵연료(nuclear spent fuel)’, ‘기술(technology)’ 등의 키워드가 다른 단어들에 비해 많이 사용되며 다양한 주제에서 이러한 키워드 관련 이슈가 다뤄졌음을 알 수 있다. 또한, 외향 연결정도 중심성(Out-Degree centrality) 결과, ‘미국(US)’, ‘파이로프로세싱(pyroprocessing)’, ‘사용후핵연료(nuclear spent fuel)’, ‘농축(enrichment)’, ‘협정(agreement)’ 등의 키워드들이 상위에 등장하는 것을 볼 수 있다.

Figure 4는 내향/외향 연결정도 중심성 값을 바탕으로 이를 도식화하였다. ‘재처리(reprocessing)’, ‘사용후핵연료(nuclear spent fuel)’, ‘협정(agreement)’ 등의 핵심 노드를 중심으로 다양한 노드들이 연결되어있는 것을 볼 수 있다. 이러한 결과를 보아 해

<Table 3> In/ Out Degree Centrality analysis in 2007 to the mid 2010's

rank	2007 - mid 2010's	
	In-Degree Centrality	Out-Degree Centrality
1	agreement (0.047619)	US (0.083333)
2	reprocessing (0.047619)	pyroprocessing (0.059524)
3	nuclear spent fuel (0.035714)	nuclear spent fuel (0.047619)
4	technology (0.035714)	enrichment (0.047619)
5	disposal (0.023810)	agreement (0.035714)



<Figure 4> Degree Centrality network in 2007 to the mid 2010's

당 시기에 사용후핵연료 건식처리기술에 대한 사회적 인식이 협정과 같은 외교적 의미가 짙었다는 것을 알 수 있다. 또한, 그래프의 좌측에 ‘원전

(nuclear power plant)’ 노드를 중심으로 ‘수출 (exportation)’, ‘공급(supply)’, ‘안정(stabilization)’ 등의 키워드가 연결되어있는 것으로 보아 원자력에

대해 사회인식이 긍정적이었음을 유추할 수 있다.

4.3.2 2010년대 중반 - 2020년 의미 네트워크 분석

<Table 4>는 2010년대 중반부터 현재까지의 사용후핵연료 처리기술 관련 기사 총 775건의 텍스트 데이터를 바탕으로 내향외향 연결정도 중심성을 분석한 결과이다. 내향 연결정도 중심성 도출결과, ‘사용후핵연료(nuclear spent fuel)’, ‘기술(technology)’, ‘안전(safety)’, ‘계획(plan)’, ‘개발(development)’ 등의 키워드가 등장하였고 외향 연결정도 중심성 결과로는 ‘연구(research)’, ‘안전(safety)’, ‘개발(development)’, ‘원전(nuclear power plant)’, ‘정부(government)’ 단어가 상위권에 도출되었다.

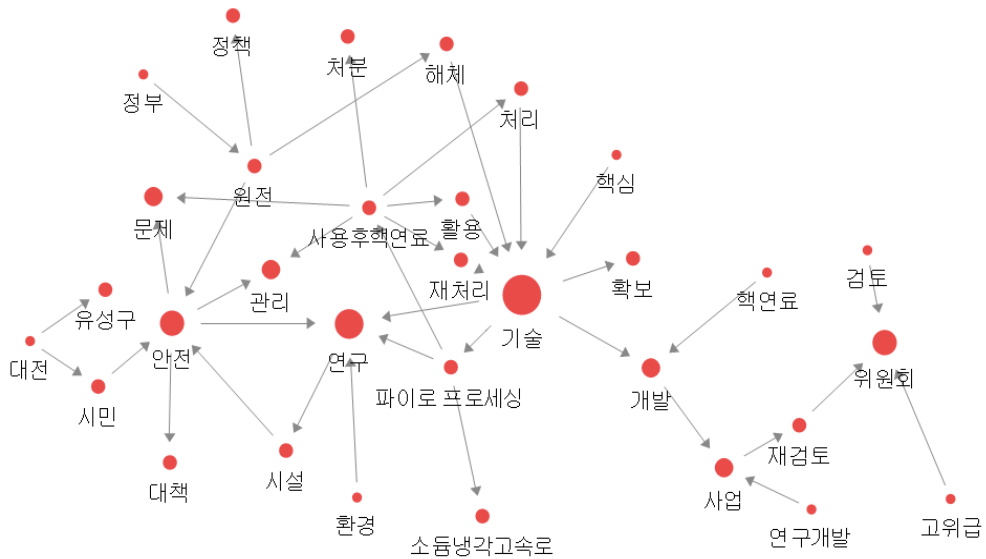
<Figure 5>는 도출된 결과를 바탕으로 연결정도 중심성을 도식화하였다. ‘기술(technology)’, ‘연구(research)’, ‘안전(safety)’, ‘위원회(committee)’ 등의 노드를 중심으로 다양한 키워드들이 연결되어있는 것을 볼 수 있다. 또한, 이전 분석결과인 <Figure 4>와 비교해보았을 때 ‘시민(citizen)’, ‘대책(countermeasure)’, ‘재검토(reexamination)’, ‘위원회(committee)’, ‘정책(policy)’ 등의 사회정치적 키워드가 새롭게 등장한 것으로 보아 원자력이 기술

분야에서 사회 분야의 이슈로 변화했음을 유추할 수 있다. ‘원전(nuclear power plant)’, ‘사용후핵연료(nuclear spent fuel)’ 키워드 관련하여서도 ‘해체(disassemble)’, ‘처분(disposal)’, ‘문제(problem)’ 등의 키워드가 함께 등장함으로써 사용후핵연료 건식처리기술에 대한 사회적 인식이 다소 고식적으로 고려되고 있음을 알 수 있다.

이러한 결과를 종합적으로 살펴본다면, 초기에는 사용후핵연료 건식처리기술에 대한 언론의 인식이 ‘미국(US)’, ‘협정(agreement)’, ‘원전(nuclear power plant)’, ‘수출(exportation)’ 등 외교적이고 긍정적인 개념으로 인식되었음을 알 수 있었다. 그러나 시간이 흐름에 따라 ‘안전(safety)’, ‘재검토(reexamination)’, ‘대책(countermeasure)’, ‘처분(disposal)’, ‘해체(disassemble)’ 등의 키워드가 출현하며 건식처리기술에 대한 지속 여부가 고려되고 있음을 알 수 있다. 정치 외교적 기술로 인식되던 사용후핵연료 건식처리기술이 국내 정책의 변화로 연구 지속 가능성이 모호해짐에 따라 사회적 인식도 점차 변화했다는 것을 함의한다. 위 연구분석 결과, 원자력 공학 기술이 정책 및 사회과학 분야와 분리하여 연구되어야 할 것이 아닌 융합적 연구의 필요성을 확인할 수 있었다.

<Table 4> In/ Out Degree Centrality analysis in the mid to late 2010's

rank	the mid to late 2010's	
	In-Degree Centrality	Out-Degree Centrality
1	nuclear spent fuel (0.041769)	research (0.048321)
2	technology (0.040950)	safety (0.041769)
3	safety (0.027846)	development (0.032760)
4	plan (0.021294)	nuclear power plant (0.031122)
5	development (0.020475)	government (0.019656)



<Figure 5> Degree Centrality network in the mid to late 2010's

5. 결론 및 시사점

본 연구는 텍스트마이닝 분석기술을 적용하여 사용후핵연료 건식처리기술에 대한 언론 인식 변화를 객관적으로 분석하였다. ‘파이로프로세싱’ 및 ‘소듐 냉각고속로’ 키워드를 포함하는 네이버 웹 뉴스 기사 전문의 텍스트데이터를 파이썬 코드를 통해 수집하여 기간에 따라 인식변화를 파악하였다. 분석 기간은 첫 기사가 게재되었던 2007년부터 2020년 현재까지로 설정하였고 빈도분석을 바탕으로 한 워드 클라우드 작성, TF-IDF 동계가중치 도출, 연결정도 중심성 산출 등의 분석방법을 통해 텍스트데이터에 대한 세부적이고 다층적인 분석을 수행하였다. 키워드의 빈도분석 결과 2010년대 중반에 사용후핵연료 건식처리기술에 대한 언론의 인식 변화가 있었음을 파악할 수 있었고 이는 2016년 발생한 경주 지진, 2017년 새 정부의 에너지 전환정책 시행 등의 영향으로 보였다. 따라서 해당 시기를 기준으

로 동향 분석이 시행되었고, 단어 빈도분석, TF-IDF, 연결정도 중심성 값 및 의미 네트워크 그래프를 도출하였다.

연구 결과, 2010년대 이전에는 사용후핵연료 건식처리기술에 대한 언론 인식이 외교적이고 긍정적이었음을 알 수 있었다. 그러나 시간이 흐름에 따라 ‘안전(safety)’, ‘재검토(reexamination)’, ‘대책(countermeasure)’, ‘처분(disposal)’, ‘해체(disassemble)’ 등의 키워드 출현빈도가 급증하며 사용후핵연료 건식처리기술연구에 대한 지속 여부가 사회적으로 진지하게 고려되고 있음을 알 수 있었다. 정치 외교적 기술로 인식되던 사용후핵연료 건식처리기술이 국내 정책의 변화로 연구 지속 가능성이 모호해짐에 따라 사회 인식도 함께 변화했다는 것을 확인하였다.

이는 곧 ‘사용후핵연료 처리기술’ 자체의 이슈보다는 원자력 정책과 같은 국내 정책 변화가 언론 인식 변화에 더 큰 영향을 끼치고 있음이라 볼 수 있

다. 원자력 정책이 사용후핵연료보다 사회적으로 더 많이 논의되고 대중에게 친숙한 주제이기 때문인 것으로 보인다. 따라서 사용후핵연료 처리기술에 대한 사회인식을 개선하기 위해서는 이와 관련한 정보를 충분히 제공해야 할 것이며, 원자력 정책 이슈와의 연계 또한 좋은 방안이 될 것이라 보인다.

그 외에도 본 연구 결과를 통해 원자력에서의 사회과학 연구에 대한 중요성이 부각되었다. 원자력 공학 분야에 사회과학 분야를 폭넓게 적용할 필요가 있으며, 국가 정책적 변화를 고려해야 원자력 산업이 지속 가능할 것임을 확인할 수 있었다.

그러나 본 연구는 사용후핵연료 건식처리기술인 '파이로프로세싱' 등과 같이 세부 연구 분야에 국한하여 빅데이터 분석방법을 적용했다는 한계가 있다. 또한, 사회 인식변화의 원인에 대한 뚜렷한 근거는 제시하지 못했으며, 뉴스 기사만을 분석하여 사회인식을 파악하려 하였다. 향후 댓글까지 고려하여 원자력 전반에 대한 언론 동향 분석연구를 진행한다면 좀 더 신뢰도 높은 결과를 도출할 수 있을 것이며, 원자력 정책연구 분야에 효율적으로 이용될 수 있을 것이라 기대된다.

최근 코로나로 인한 뉴노멀 시대의 도래와 같은 사회적 분위기를 계기로 인공지능, 빅데이터 연구와 같은 언택트 관련 기술의 발전이 더욱 가속화되고 있다. 각종 연구 분야에서는 이러한 연구 트렌드를 따라가기 위한 융합적 연구가 본격적으로 시행되고 있으나 원자력 분야의 경우 자연어 처리, 텍스트마이닝 분석 등 인공지능 및 빅데이터 관련 기술을 적용한 연구는 많이 수행되지 않았다. 이에 본 연구가 갖는 학술적 의의는 원자력 연구 분야에 데이터 사이언스 분석기술의 적용 가능성을 확인할 수 있었다는 것이다. 또한, 현 정부의 원전 감축과 같은 에너지 정책의 영향으로, 사용후핵연료 처리 기술 연구개발에 대한 재평가가 시행되는 이 시점

에서 해당 분야의 주요 키워드 분석은 향후 연구 방향 설정에 기여할 수 있을 것이라는 측면에서 실무적 의의를 갖는다. 원자력에 대한 안전 기술 및 공학적 무결성에만 집중할 것이 아닌 외부 타인들의 시선도 적절히 고려해야 할 것이며, 더 나아가 원자력 공학 기술을 내부적으로만 논의하는 것이 적절할지도 재고하여 반문해봐야 할 것이다. 이 외에도 원자력 관련 다각도의 범분야적 연구가 수행된다면 원자력 산업 유지를 위한 합리적인 대안을 마련할 수 있을 것이다.

참고문헌(References)

- Chakraborty, G., Pagolu, M., & Garla, S. Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute.(2014).
- Charlton, William S., et al, 'Proliferation resistance assessment methodology for nuclear fuel cycles.', Nuclear Technology 157.2 (2007), 143-156.
- Chebeskov. A., et al, 'The GIF Proliferation Resistance and Physical Protection (PR&PP) Evaluation Methodology: Status, Applications and Outlook'. (2017).
- D.R.Kim, 「Knowledge Structure Analysis on Business Administration Research using Keyword Network Analysis」, Ph.D Degree Thesis, Kyungil.University. (2016).
- D.S.Kim, 「A Study on the Lexicon Development for Public Opinion Trend Analysis on Social Media : A Case Study of Twitter Opinion on Nuclear Power」, Master Degree Thesis, Hanyang University. (2015).

- E.S.Yoo·G.H.Choi·S.H.Kim, 'Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels', 『Journal of The Korea Society of Computer and Information』, vol 20-2.(2015).
- Freeman, L. C., Roeder, D., & Mulholland, R. R. 「Centrality in social networks: II. Experimental results. Social networks, 2(2), (1979), 119-141」 .
- Generation IV International Forum, 'A Technology Roadmap for Generation IV Nuclear Energy Systems' Tech. Rep, (2002).
- Generation IV International Forum PRPP Working Group and System Steering Committees, 'Proliferation Resistance and Physical Protection of the Six Generation IV Nuclear Energy Systems Tech. Rep'. (2011).
- H.D. Kim, S.K. Ahn, and J.S. Jeong, 'Evaluation Methodology for Proliferation Resistance and Protection of Generation IV Nuclear Energy Systems', KAERI/TS-333/2018. (2018).
- IDC Digital Universe Study, 'extracting value from chaos', <http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm> (2011).
- Internet trend, <http://www.internettrend.co.kr/trend/Forward.jsp>, (2020).
- J.G.Ko, Y.S.Bae, J.Y.Park and K.Park, 'Technologies Trends in Image Big Data Analysis', 『2014 Electronics and Telecommunications Trends, 29(4), (2014).
- J.H.Ahn, 『Text Network Analysis of Shin-Kori Unit 5-6 Deliberative Polling』, Master Degree Thesis, Sejong University. (2020).
- J.H.Han. 『Social Network Analysis on Mapping the Knowledge Structure of Dementia Research』, Ph.D Degree Thesis, Chungang University. (2015).
- John Gantz & David Reinsel, 'Extracting Value from Chaos', IDC Digital University Study IVIEW June. (2011).
- J.S.Jeong, 『A Study on the Research Environment Change of Proliferation Resistance and Social Recognition Trends in the Spent Nuclear Fuel Treatment Technology Using Big Data Analysis Method』, Master Degree Thesis, University of science and technology. (2021).
- J.Y.Jang, 'A Study on Research Trends of Graph-Based Text Representations for Text Mining', 『The Journal of The Institute of Internet, Broadcasting and Communication 13.5 (2013), 37-47』 .
- K.H.Eom, 『Python for Political Science』. Kyungpook National University Press, (2018), pp.147-152
- Kim, J. H., & Lee, J. M, 『Comparison and analysis of domestic and foreign sports brands using text mining and opinion mining analysis』. The Journal of the Korea Contents Association, 18(6), (2018), 217-234.
- Lambert, H, Elayat, H. A., Szytel, L., & Dreicer, M, 『LISSAT Analysis of a Generic Centrifuge Enrichment Plant (No. UCRL-CONF-23188 1)』, Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), (2007).
- Lim, D. H, 『big data analysis using R』. Paju: Free Academy, (2015).
- M.K.Kim, 'September 12, 2016 Gyeongju earthquake and nuclear safety', 『Korean Society of Civil Engineers』, 65(4), (2017), 31-35.
- N. J.Van Eck, L. Waltman, Software survey: VOSviewer, 'a computer program for bibliometric mapping', Scientometrics 84,

- (2009), 523-538.
- N. J. Van Eck, L. Waltman, 'Text mining and visualization using VOSviewer', arXiv preprint arXiv:1109.2058 (2011).
- N.J.Van Eck, Waltman, L, 'Text mining and visualization using VOSviewer', ArXiv, abs/1109.2058.Ahn, H., K.-j. Kim, and I. Han, "Purchase Prediction Model using the Support Vector Machine," Journal of Intelligence and Information Systems, Vol.11, No.3(2005), 69-81.
- N. J. van Eck, L. Waltman, 'Citation-based clustering of publications using CitNetExplorer and VOSviewer', Scientometrics 111, (2017), 1053-1070.
- Park, E. L., & Cho, S, 'KoNLPy: Korean natural language processing in Python', Annual Conference on Human and Language Technology, (2014), 133-136, Human and Language Technology.
- Salton, G., Wong, A., & Yang, C. S, A vector space model for automatic indexing. Communications of the ACM, 18(11), (1975). 613-620.
- S.K. Ahn et al, 'A Proliferation Resistance Evaluation for a Pyroprocessing Facility Design', INMM 55th Annual Meeting, (2014).
- S.S.Lee, 『Network Analysis Methodology』, Nonhyun, (2016).
- Statistics Training Institute, 'How text is analyzed', http://sti.kostat.go.kr/window/2017b/html/2017_win_3.html. (2017).
- S.Y.Kim, Y.J.Chung, 『처음 배우는 머신러닝』, Hanbit Media INC, (2017).
- Van Atteveldt, W. H. 『Semantic network analysis: Techniques for extracting, representing, and querying media content』. Charleston, SC: BookSurgePublishers.(2008).
- Y.H.Kim, 『Social Network Theory』. Parkyeongsa. (2008).
- Y.H.Kim, Y.J.Kim, 『Social Network Analysis』. Parkyeongsa, (2016).
- Y.S.Park, 『Base on Machine Learning, prediction threat index of events about incident's attack』, Master Degree Thesis, Graduate School Korea University. (2018).

Abstract

Analysis of media trends related to spent nuclear fuel treatment technology using text mining techniques

Ji-Song Jeong* · Ho-Dong Kim**

With the fourth industrial revolution and the arrival of the New Normal era due to Corona, the importance of Non-contact technologies such as artificial intelligence and big data research has been increasing. Convergent research is being conducted in earnest to keep up with these research trends, but not many studies have been conducted in the area of nuclear research using artificial intelligence and big data-related technologies such as natural language processing and text mining analysis. This study was conducted to confirm the applicability of data science analysis techniques to the field of nuclear research. Furthermore, the study of identifying trends in nuclear spent fuel recognition is critical in terms of being able to determine directions to nuclear industry policies and respond in advance to changes in industrial policies. For those reasons, this study conducted a media trend analysis of pyroprocessing, a spent nuclear fuel treatment technology. We objectively analyze changes in media perception of spent nuclear fuel dry treatment techniques by applying text mining analysis techniques. Text data specializing in Naver's web news articles, including the keywords "Pyroprocessing" and "Sodium Cooled Reactor," were collected through Python code to identify changes in perception over time. The analysis period was set from 2007 to 2020, when the first article was published, and detailed and multi-layered analysis of text data was carried out through analysis methods such as word cloud writing based on frequency analysis, TF-IDF and degree centrality calculation. Analysis of the frequency of the keyword showed that there was a change in media perception of spent nuclear fuel dry treatment technology in the mid-2010s, which was influenced by the Gyeongju earthquake in 2016 and the implementation of the new government's energy conversion policy in 2017. Therefore, trend analysis was conducted based on the corresponding time period, and word frequency analysis, TF-IDF, degree centrality values, and semantic network graphs were derived.

* Master candidate, Department of Quantum Energy and Chemical Engineering, University of Science and Technology

** Corresponding Author: Hodong Kim

Korea Atomic Energy Research Institute

Daedeok Daero 989, Yuseong-gu, Daejeon 34507, Korea

E-mail: khd@kaeri.re.kr

Studies show that before the 2010s, media perception of spent nuclear fuel dry treatment technology was diplomatic and positive. However, over time, the frequency of keywords such as "safety", "reexamination", "disposal", and "disassembly" has increased, indicating that the sustainability of spent nuclear fuel dry treatment technology is being seriously considered. It was confirmed that social awareness also changed as spent nuclear fuel dry treatment technology, which was recognized as a political and diplomatic technology, became ambiguous due to changes in domestic policy.

This means that domestic policy changes such as nuclear power policy have a greater impact on media perceptions than issues of "spent nuclear fuel processing technology" itself. This seems to be because nuclear policy is a socially more discussed and public-friendly topic than spent nuclear fuel. Therefore, in order to improve social awareness of spent nuclear fuel processing technology, it would be necessary to provide sufficient information about this, and linking it to nuclear policy issues would also be a good idea. In addition, the study highlighted the importance of social science research in nuclear power. It is necessary to apply the social sciences sector widely to the nuclear engineering sector, and considering national policy changes, we could confirm that the nuclear industry would be sustainable.

However, this study has limitations that it has applied big data analysis methods only to detailed research areas such as "Pyroprocessing," a spent nuclear fuel dry processing technology. Furthermore, there was no clear basis for the cause of the change in social perception, and only news articles were analyzed to determine social perception. Considering future comments, it is expected that more reliable results will be produced and efficiently used in the field of nuclear policy research if a media trend analysis study on nuclear power is conducted.

Recently, the development of uncontact-related technologies such as artificial intelligence and big data research is accelerating in the wake of the recent arrival of the New Normal era caused by corona. Convergence research is being conducted in earnest in various research fields to follow these research trends, but not many studies have been conducted in the nuclear field with artificial intelligence and big data-related technologies such as natural language processing and text mining analysis. The academic significance of this study is that it was possible to confirm the applicability of data science analysis technology in the field of nuclear research. Furthermore, due to the impact of current government energy policies such as nuclear power plant reductions, re-evaluation of spent fuel treatment technology research is undertaken, and key keyword analysis in the field can contribute to future research orientation. It is important to consider the views of others outside, not just the safety technology and engineering integrity of nuclear power, and further reconsider whether it is appropriate to discuss nuclear engineering technology internally. In addition, if multidisciplinary research on nuclear power is carried out, reasonable alternatives can be prepared to maintain the nuclear industry.

Key Words : Nuclear spent fuel, text mining, word cloud, TF-IDF, Semantic Network Analysis

Received : February 8, 2021 Revised : June 1, 2020 Accepted : June 24, 2020

Corresponding Author : Hodong Kim

저자 소개



정지송

과학기술연합대학원대학교(UST) 한국원자력연구원(KAERI) 캠퍼스에서 양자 에너지화학공학과 석사과정을 졸업하였고, 주요 관심 연구 분야는 텍스트마이닝과 의미 네트워크 분석 등이다.



김호동

현재 한국원자력연구원 책임연구원과 UST 교수로 재직 중이다. 고려대학교 화학공학과에서 박사학위를 취득하였고 미국 로스앨라모스 국립연구소에서 박사후연수과정을 마쳤다. 현재 한국핵물질관리학회 부회장을 역임 중이다. 주요 관심분야는 원자력 시스템의 핵비확산 및 안전조치 데이터 분석, 핵물질 관리 방법론 개발 등이다.