

특허 텍스트마이닝을 활용한 바이오 연료 기술 모니터링*

이다혜** · 최하영*** · 정병기**** · 윤장혁*****

I. 서 론

II. 바이오 연료

III. 분석 절차

1. 데이터 수집, 전처리, 키워드 추출
2. 토픽모델링(Topic modeling)
3. 네트워크 분석(Network analysis)
4. 정보 엔트로피(Information entropy)

IV. 분석 결과

1. 바이오 연료 특허 동향
2. 세부 기술요소 정의
3. 세부 기술 간 네트워크 생성 및 분석
4. 기술요소별 정보 엔트로피 분석

V. 결론 및 추후연구

* 본 논문은 2017년 제12회 대학(원)생 지식재산 우수논문 공모전의 수상작을 개선 및 발전시킨 논문임. 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2015R1A1A1A05027889).

** 건국대학교 산업공학과(제1저자).

*** 건국대학교 산업공학과.

**** 건국대학교 산업공학과.

***** 건국대학교 산업공학과 부교수(교신저자).

초 록

심각한 환경오염의 위협으로 인해 전 세계적으로 신재생 에너지 확충 노력이 지속되고 있는 가운데, 바이오 연료는 친환경성과 활용성 측면에서의 장점으로 큰 각광을 받고 있는 신재생 에너지이다. 바이오 연료에 대한 큰 관심에 비해, 바이오 연료의 세부 기술에 대한 정의와 이를 통한 기술 모니터링에 대한 계량분석연구는 미흡한 실정이다. 따라서 본 연구는 바이오 연료에 관한 특허 자료를 바탕으로 세부 기술들의 동향 및 특성을 분석한다. 본 연구의 분석 과정은 1) 바이오 연료와 관련된 특허 전문을 수집하여 바이오 연료 세부 기술과 관련된 키워드를 추출한 뒤, 2) 특허 문서와 키워드 집합을 기반으로 토픽모델링 기법을 실시하여 바이오 연료의 세부 기술을 정의하고, 3) 정의된 세부 기술의 전반적 동향을 살펴보기 위하여 각 세부 기술의 연도구간별 동시 출현 네트워크를 구축하여 기술 동향을 분석한 후, 4) 정보 엔트로피 분석을 통하여 각 세부 기술의 연결강도 및 다양성을 살펴본다. 본 분석과정을 통하여 바이오 연료의 세부 기술을 파악하고 기술 간 상호관계 및 변화 추세를 확인하였다. 분석결과로는 바이오 연료의 원천 기술이 개발된 후 유전공학 분야를 응용하여 발전하였으며, 향후 미세조류와 관련된 기술 개발이 지속될 것으로 전망된다. 본 연구의 분석결과는 향후 바이오 에너지 분야의 기업들과 연구자들이 기술의 변화과정을 파악하고 그에 따른 기술개발의 방향을 설정하는 데 활용될 수 있을 것이다.

주제어

바이오 연료, 기술 모니터링, 텍스트 마이닝, 토픽모델링, 네트워크 분석, 정보 엔트로피

I. 서 론

전 세계적으로 화석연료 수요의 급증으로 인한 막대한 양의 매연, 오수 등 환경오염이 발생하고 있으며, 최근에는 고유가, 화석연료 고갈 문제와 함께 국가 경쟁력 제고를 위한 신성장 동력산업인 신재생 에너지 확충 노력이 절실해지고 있다. 친환경 에너지 가운데 바이오 에너지는 전기 생산 및 난방용과 수송용 등 다양한 용도로 활용될 수 있는 잠재성과 원료의 풍부함, 지속가능성으로 인해 화석연료의 대체 자원으로 급부상하고 있으며, 현재 여러 국가와 기업들의 관심이 바이오 연료 기술 개발에 집중되고 있다.

바이오 연료에 관한 기술적 관심은 증가하고 있는 반면, 계량화된 자료 분석 과정을 통하여 세부기술들을 심도 있게 분석한 연구는 미흡한 실정이다. 기존 연구들은 정성적 분석을 통해 해당 기술의 정책개발 및 산업동향을 다루거나, 경제성, 타당성 분석에서도 간단한 계량적 지표를 활용하여 분석을 진행하였다.¹⁾²⁾ 이러한 분석은 변화하는 세부기술과 이들 간의 상호관계를 파악하는 데는 한계점을 지니고 있다. 따라서 본 연구에서는 바이오 연료에 관련된 특허문서를 수집 및 분석하여 바이오 연료를 구성하는 세부 기술 요소들을 파악하고 이들의 발전 동향을 분석하는 것을 목표로 한다. 특허문서는 산업상 이용가능성이 있는 응용기술 내용을 포함하는 신뢰성 있는 정보이며 기술동향의 분석 및 기술전략의 수립에 필수적인 요소로 인식되어 왔다.³⁾ 따라서 특허를 활용한 분석은 바이오 연료 기술을 다루는 기업과 연구자들에게 해당 기술에 대한 변화하는 동향 정보를 제공하며 이에 따른 향후

1) 민은주 · 김수덕, “수송용 바이오에탄올 도입의 경제성 평가 및 Cdm 사업 타당성 분석”, 『에너지경제연구』, 제7권 제1호(2008), 133-154면.

2) 송효학 · 박종명, “바이오에너지 산업의 현황 및 최신 연구동향”, 『BT NEWS』, 제20권 제2호(2013), 22-25면.

3) Yoon Byungun & Park Yongtae, “A Text-Mining-Based Patent Network: Analytical Tool for High-Technology Trend”, *The Journal of High Technology Management Research*, Vol.15 No.1(2004), pp.37-50.

연구 방향 설정에 도움이 될 수 있다.

본 연구에서는, 바이오 연료 기술관련 IPC 범위에 해당되는 USPTO의 등록 특허를 사용하여 분석을 진행한다. 분석을 위해, 먼저 수집된 특허의 전문에 대해 키워드를 추출하고, 키워드 집합과 빈도를 기반으로 토픽모델링(Topic modeling) 기법을 활용하여 바이오 연료의 세부 기술을 도출한다. 이들 세부 기술의 전반적인 변화를 모니터링하기 위하여, 세부 기술들을 네트워크로 시각화하여 전반적인 기술 동향을 파악한다. 다음으로, 각 세부기술의 상호관계 분석을 위하여 정보 엔트로피(Information entropy) 분석을 진행한다. 토픽모델링 기법은 각 특허 문서가 지니는 특정 주제를 파악할 수 있으며 더 나아가 해당 주제를 구성하는 키워드 분포를 도출할 수 있으므로, 바이오 연료와 관련된 세부기술의 정의가 가능하다. 또한 네트워크 분석은 전체 기술 내 각 세부 기술별 연결강도와 연도구간별 핵심기술 파악 및 기술 간 상호관계 등을 살펴볼 수 있으며, 이를 통해 각 기술의 모니터링 분석을 실시할 수 있다. 정보 엔트로피 분석은 다른 기술들과 연결된 해당 기술의 연관다양성 분석 및 향후 발전가능성 분석에 활용된다.

본 연구의 분석결과는 바이오 연료 분야의 기업과 연구자들에게 기술 발전 흐름 및 동향 정보를 제공하며, 이에 따른 향후 기술개발의 방향을 설정하거나 바이오 연료 분야의 발전 동향에 대한 기초 연구 자료로써 활용될 수 있을 것이다.

본 논문의 구성으로는 제2장에서 바이오 연료에 관하여 설명하고, 제3장에서는 연구의 분석절차와 활용되는 이론적 배경연구를 설명한다. 제4장에서는 분석절차에 따른 분석결과를 제시하며 마지막 제5장에서는 연구의 결론, 한계점 및 추후연구에 대해 기술한다.

II. 바이오 연료

바이오 연료란 태양광을 이용하여 광합성 되는 유기물 및 동 유기물을 소

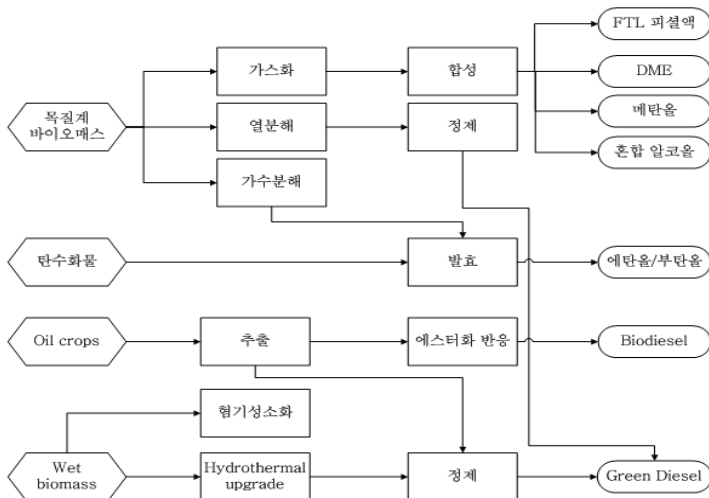
비하여 생성되는 모든 생물 유기체(바이오매스)의 에너지를 의미한다.⁴⁾ 이러한 바이오매스는 풍부하고 친환경적이며 탄소 중립적이라는 점에서 온실가스 저감에 탁월하다는 장점을 지니고 있다. 바이오 연료의 분류에 관한 정확한 기술적 정의는 없지만, 원료를 기준으로 1세대, 2세대 그리고 3세대로 나뉘게 된다. 이러한 분류로 그림 1과 같이 각 원료마다 다른 가공과정을 거쳐 여러 가지 바이오 연료로 생성된다. 바이오 연료의 용도로는 크게 전기 생산/난방용과 수송용으로 구분되며 각각 용도로 활용될 수 있는 연료의 형태로는 바이오에탄올, 바이오부탄올, 바이오디젤, Fischer-Tropsch 액(FTL), 바이오가스 등이 존재한다.

우선, 1세대 연료는 식물에 의해 생산되는 바이오매스의 식용 가능한 특정부분을 사용하며, 상대적으로 단순한 가공 과정을 거쳐 연료가 생산된다. 1세대 연료로는 옥수수, 감자에서 추출한 전분과 사탕수수, 사탕무에서 추출한 설탕을 발효하여 만든 바이오 에탄올, 부탄올이 있으며, 오일 시드 작물의 에스테르 교환 반응에 의해 제조되는 바이오 디젤 또한 널리 알려져 있다. 1세대 연료는 가장 잘 알려진 공급 원료와 간단한 생산 방법으로 인해 상용화되었으나 식용 가능한 바이오매스라는 점에서 식량과의 경쟁으로 인한 곡물가격 상승 등으로 경제성에 한계를 지니고 있다. 이러한 식품 대 연료 경쟁을 제한하기 위하여 2세대 연료는 목질 및 볏짚, 식량작물 생산의 비식용 잔여물 등을 원료로 사용한다. 이러한 원료를 2세대 연료로 전환하는 프로세스는 두 가지로 분류되는데, 첫 번째로는 2세대 에탄올, 부탄올을 생산하는 생화학적 공정으로 기본적인 생산 단계는 전처리, 당화, 발효 및 증류를 포함한다. 다음은 열화학적 공정으로 생화학 공정보다 높은 압력과 온도를 가하여 생산하게 되며, 가스화 또는 열분해를 거쳐 합성 및 정제과정을 통해 최종 연료인 메탄올, Fischer-Tropsch 액(FTL), dimethyl ether(DME), 혼합 알코올 등을 생산한다. 1세대, 2세대에 이어 미래의 3세대 바이오매스로 미세조류가 급부상하고 있다. 미세조류는 태양 에너지로부터 광합성을

4) 안두현 외 3인, “바이오에너지 연구개발 동향과 시사점”, 과학기술정책연구원, 2007, 1-137면.

하여 이산화탄소를 고정하는 해양 생명체로서 고정된 이산화탄소로부터 바이오 연료, 음식물 등의 다양한 유용 물질을 만들어 낼 수 있다.⁵⁾ 미세조류는 탈수과정과 조류대사산물의 추출 및 정제를 거친 후 연료로의 전환 단계가 진행된다. 전환기술로는 2세대 연료와 마찬가지로 생화학적, 열화학적 공정으로 나뉘며 생화학적 전환은 유기폐기물을 바이오가스로 변환시키는 혐기성소화와 알코올발효, 광 생물학적 수소생산 등을 포함하며, 열화학적 전환은 직접연소, 가스화, 열분해 등 온도와 압력을 가하여 연료로 전환된다.

〈그림 1〉 바이오 연료 제조 과정



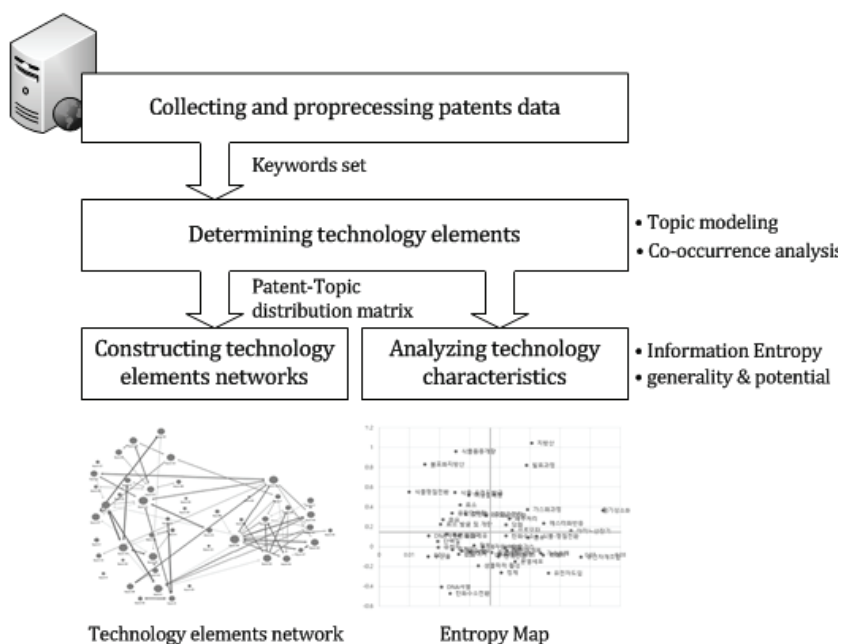
III. 분석 절차

본 연구가 진행되는 분석 과정은 그림 2의 절차를 따른다. 1) 먼저, 바이오 연료 관련 특허 문서를 수집하여 전처리 과정을 진행한다. 2) 다음으로 토픽 모델링 기법을 활용하여 바이오 연료 기술을 구성하고 있는 세부 기술요소

5) Schenk et al., "Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production", *Bioenergy Research*, Vol.1 No.1(2008), pp.20-43.

들을 정의한다. 3) 특허 문서를 연도구간별로 분류하여 각 세부 기술요소들의 동시 출현 매트릭스를 구성한 후 네트워크 분석을 진행하여 핵심 세부기술 및 기술 간 상호관계를 파악한다. 마지막으로 4) 동시출현 매트릭스를 활용하여 세부 기술들의 정보 엔트로피 분석을 통해 기술들의 다양성 및 발전 가능성과 향후 발전가능성을 파악한다.

〈그림 2〉 분석 절차



1. 데이터 수집, 전처리, 키워드 추출

본 연구는 바이오 연료의 세부 기술 동향을 분석하기 위하여 특허 문서를 활용한다. 세계지적재산권기구에서 유엔 기후변화협약에 명시된 ‘IPC Green Inventory’ 범주에 있는 미국 특허청 특허를 대상으로 총 34,237건의 바이오 연료 관련 유효특허를 수집하여 분석을 진행하며, 기술적 유의성을 판단하기 어려운 단순 출원 상태의 특허를 분석대상에서 제외한다.

다음으로, 총 34,237개의 수집된 특허 자료 전문을 활용하여 기술적으로 의미가 있는 바이오 연료 관련 키워드와 키워드의 빈도를 추출한다. 키워드 추출을 위해 본 연구에서는 인간의 학습 능력과 같은 기능을 컴퓨터에서 실현한 기법인 머신러닝 알고리즘을 기반으로 자연어를 처리하고 분석하는 Alchemy API 웹 서비스를 활용한다(<https://www.ibm.com/watson/alchemy-api.html>). 따라서 해당 웹 서비스를 활용하여 바이오 연료 분야의 특허 문서로부터 키워드를 추출하며, 전처리 과정을 거쳐 총 21,274개의 유효 키워드 집합을 구성한다.

초기 키워드 집합에서는 바이오 연료와 기술적으로 관련 없는 불용어를 포함하기 때문에 전처리 과정을 통하여 기술적 내용을 포함하지 않는 키워드, 특수기호를 포함하는 키워드를 제거한다. 또한 세부 기술의 명확한 기술적 해석을 위하여 두 단어 이상으로 구성된 키워드와 총 단어의 길이가 3 이상, 35 이하인 키워드만 선별하여 21,274개의 최종 유효 키워드 집합을 구성하였다. 이러한 과정을 통해 각 유효 특허문서를 기술적 의미를 지닌 키워드 벡터로 변환하여 특허문서를 구조화한 후 바이오 연료에 대한 세부 기술을 도출하기 위하여 다음 단계인 토픽모델링에 활용한다.

2. 토픽모델링(Topic modeling)

토픽모델링은 문서 집합을 구성하는 단어를 분석함으로써 주제를 추론하는 확률적 모형이다. 모든 문서는 단어들의 벡터로 표현될 수 있으며, 이러한 벡터들의 조합은 문서의 주제를 나타낸다. 토픽모델링은 구조화되지 않은 방대한 문서 집단에서 잠재적 주제를 찾아내고 대량 문서를 분류하기 위한 기법으로 활용될 수 있어 다양한 연구들이 진행되어 왔다.⁶⁾⁷⁾ 본 연구에

6) Jeong Byeongki & Yoon Janghyeok, "Competitive Intelligence Analysis of Augmented Reality Technology Using Patent Information", *Sustainability*, Vol.9 No.4(2017), pp.497.

7) Yoon, Janghyeok et al., "Identifying Product Opportunities Using Collaborative Filtering-Based Patent Analysis", *Computers & Industrial Engineering*, Vol.107(2017),

서는 특허 문서의 텍스트 자료에서 세부 기술 주제를 식별하기 위하여 다양한 토픽모델링 알고리즘 중 기존 기법을 수정 보완한 가장 발전된 토픽모델링 기법으로 알려진 LDA 모델을 활용하여 분석을 진행한다. LDA는 사람이 문서를 작성한다는 가정 하에서 잠재적인 주제를 찾아내는 확률기반 생성 모형(Generative model)이며,⁸⁾ 대량의 데이터 집합에서 잠재적인 주제를 감지하는 데 유용한 텍스트 마이닝 기법이다.

〈그림 3〉 LDA 결과별 분석방법

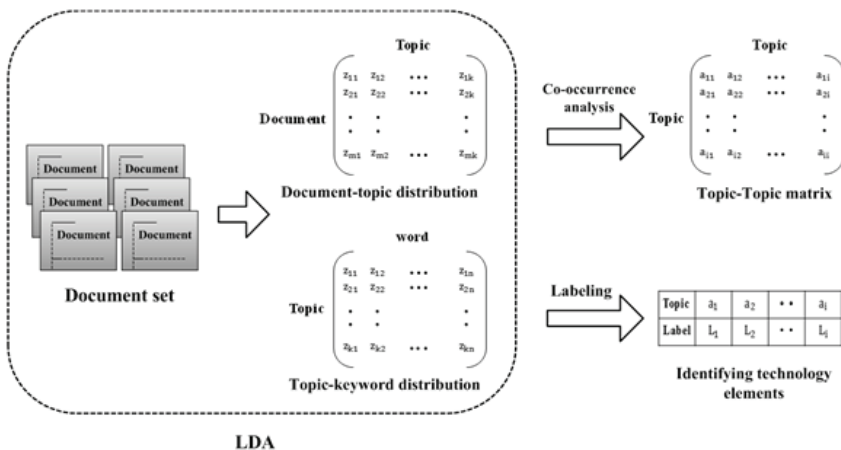


그림 3과 같이 LDA 알고리즘의 결과로는 전체 문서집합에 대한 각 문서의 토픽 분포를 나타내는 특허-토픽 매트릭스와 토픽별 구성 단어의 분포를 나타내는 토픽-키워드 매트릭스가 존재한다. 특허-토픽 매트릭스는 네트워크 분석을 진행하기 위하여 토픽들의 동시 출현 매트릭스를 구축하는데 활용되며, 토픽-키워드 매트릭스는 최적 토픽 수를 결정하고 해당 토픽의 높은 구성 확률을 가지는 상위 키워드를 확인하여 세부 구성 기술을 정의하기 위하여 활용될 것이다. 이러한 과정을 통하여 주어진 대량 문서들에 대한 문서별

pp.376-387.

8) Blei et al., "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol.3 No. Jan(2003), pp.993-1022.

토픽분포와 토픽별 단어분포를 분석함으로써 개별 문서가 다루고 있는 주제를 예측할 수 있다.

3. 네트워크 분석(Network analysis)

네트워크는 기본적으로 행위자, 개체를 나타내는 노드(node)와 노드들 간의 관계인 링크(link)로 구성된다. 네트워크를 활용하여 구성요소 간 전체 및 부분적 패턴과 구성요소 간 연결 관계를 설명하는 네트워크 분석이 진행된다. 이러한 네트워크 분석은 기본 데이터 자체로는 파악하기 어려운 구성요소 간 연결 관계를 시각적으로 나타내어 데이터 구조를 한눈에 파악할 수 있도록 한다. 또한 다양한 계량적 지표를 활용할 수 있다는 점에서 네트워크 분석을 이용하여 정보를 분석하는 연구가 활발히 진행되어 왔다.⁹⁾¹⁰⁾

분석에는 네트워크 내 링크가 연결되어 있는 패턴을 통해 네트워크 특성을 파악하는 기본 속성 분석과 한 노드가 전체 네트워크 내 중심에 위치하는 정도를 나타내는 중심성 분석으로 나뉜다. 기본 속성 분석에는 네트워크 크기, 연결정도, 밀도 등이 있으며, 중심성 지표로는 연결정도 중심성, 근접 중심성, 매개 중심성 등이 존재한다.

네트워크 분석을 통하여 각 노드와 노드 간 상호관계를 단순히 데이터로 확인하는 것이 아닌 시각적으로 구성함으로써 전체적 특징을 파악할 수 있으며 객관적 지표들을 활용하여 전체 네트워크 내 노드의 특징의 분석을 진행하게 된다.

9) Yoon Byungun & Park Yongtae, "A Text-Mining-Based Patent Network: Analytical Tool for High-Technology Trend", *The Journal of High Technology Management Research*, Vol.15 No.1(2004), pp.37-50.

10) 서원철 외 2인, "단어동시출현분석을 통한 한국의 국가 R&D 연구동향에 관한 탐색적 연구", 『Journal of Information Technology Applications & Management』, 제19권 제4호(2012), 1-18면.

4. 정보 엔트로피(Information entropy)

정보 엔트로피는 Claude E. Shannon에 의해 1948년 도에 정보이론의 일부로 처음 제시된 것으로, 정보이론에서 엔트로피란 무질서도를 의미하며 불확실성, 다양성을 나타낸다. 정보이론에서 이러한 데이터의 엔트로피를 확률적으로 산출하기 위하여 정보량이 사용되며 정보량이란 특정한 사건에 대해 모르는 정도의 양을 의미한다.¹¹⁾ 정보 엔트로피(H)는 여러 종류의 요소로 구성되어 있는 군집에서 k번째 요소의 군집 내 비율(확률)로 나타나며 식 1과 같이 계산된다.

$$H = - \sum_{i=1}^n P_i \log P_i \quad (\text{식 1})$$

식 1에 따르면 전체 데이터에서 각각의 한 요소가 등장할 확률이 비슷할수록 즉, 요소들이 균등하게 구성되어 있을수록 엔트로피는 높아지며 불확실성이 높은 것으로 해석할 수 있다. 반대로 비균등(Non-Uniform) 분포는 균등 분포보다 낮은 정보 엔트로피를 가지게 된다. 정보 엔트로피는 전자 공학 및 데이터 네트워크 분야에서 널리 활용되며 사회학 분야에서도 다양한 연구가 진행되어 왔다.¹²⁾¹³⁾ 본 연구에서는 토픽별로 연도구간에 따른 엔트로피의 절대 평균과 증가율의 기하평균을 분석함으로써 각 토픽의 인접한 토픽들에 대해 지니는 다양성과 활용성을 파악하며 해당 토픽의 다른 기술 분야와의 연계적 측면에서 향후 발전가능성을 측정하는 지표로 사용된다.

11) 정병기 외 2인, “융합기술의동향분석을 위한 의미론적 특허분석 접근 방법”, 『지식재산연구』, 제11권 제4호(2016), 211-240면.

12) 이정, “정보이론의 엔트로피 관점에서의 바라본 온라인 소비자 리뷰의 소비자 의사결정에 있어 불확실성 감소 효과”, 『한국전자거래학회지』, 제16권 제4호(2011), 241-256면.

13) 김갑수 외 2인, “엔트로피를 이용한 객체 지향 프로그램의 복잡도 척도”, 『정보과학회논문지』, 제22권 제12호(1995), 1656-1666면.

IV. 분석 결과

1. 바이오 연료 특허 동향

바이오 연료 관련 세부 기술의 동향을 파악하기 위하여 USPTO에 저장된 등록특허 가운데 세계지적재산권기구에서 유엔 기후변화협약에 명시된 'IPC Green Inventory' 내 바이오 연료 기술관련 IPC 범위를 활용하여 분석을 진행하였다. 해당 IPC 범위는 표 1과 같이 고체, 액체, 가스 연료와 유전공학 분야로 총 4가지의 분야로 나뉘게 된다. 해당 IPC 범위 내에서 등록 특허를 수집한 결과, 등록 특허의 출원일을 기준으로 1958년에 처음 등장하여 2015년까지 총 34,237건의 바이오 연료 관련 등록 특허가 출원되었으며 연도별 등록 특허의 수는 그림 4과 같다.

바이오 연료 관련 등록 특허의 전반적인 경향을 살펴보면, 액체 연료와 관련된 연구개발의 관심 증가로 인하여 1973년 등록 특허 수가 41건에서 1974년 219건으로 약 5배 급증한 것을 볼 수 있다. 그러나 바이오매스의 생산 등 기술적 한계로 인하여 1990년대 중반까지 바이오 연료 관련 등록 특허 수의 발전이 없었으나 유전공학 및 합성생물학의 발전으로 인하여 2000년대 후반 바이오 연료 관련 연구개발이 재조명되어 액체 연료의 등록 특허 건수가 급증하는 모습을 파악할 수 있다. 그 후 수송용 연료로 사용가능한 점에서 바이오 에너지 산업이 주목받아 등록 특허 건수가 급증하는 모습을 파악할 수 있다. 본 연구에서는 분석의 신뢰성을 위하여 등록 특허의 출원일을 기준으로 특허 문서를 수집하였기 때문에 2012년 이후 등록 특허 수가 줄어드는 모습을 확인할 수 있다.

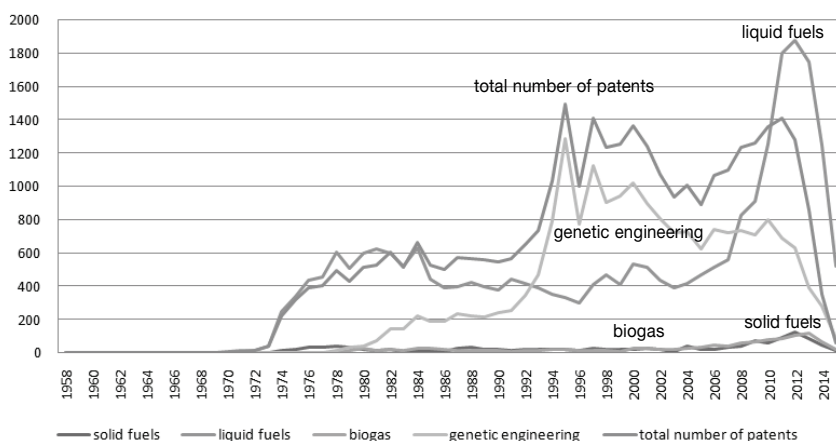
그림 5와 같이 총 34,237건의 바이오 연료 관련 등록 특허 가운데 액체 연료와 관련된 등록 특허는 25,613건이 포함되며, 4가지 분야 중 74.81%로 가장 높은 수준의 비율을 차지하고 있다. 다음으로 유기체의 유전 조작과 관련된 특허가 19,379건으로 전체의 56.60%를 차지하며, 바이오 연료 관련 기술

을 개발하는 데 있어 유전공학 분야가 널리 활용되고 있음을 확인할 수 있다. 가장 많은 비중을 차지하는 액체 연료 내에서는 바이오 디젤 관련 등록 특허가 22,554건으로 88.06%를 차지하는 것을 보아 바이오 연료 산업은 주로 바이오 디젤 생산을 지향하며 이와 관련된 연구개발이 활발히 진행될 것으로 판단된다.

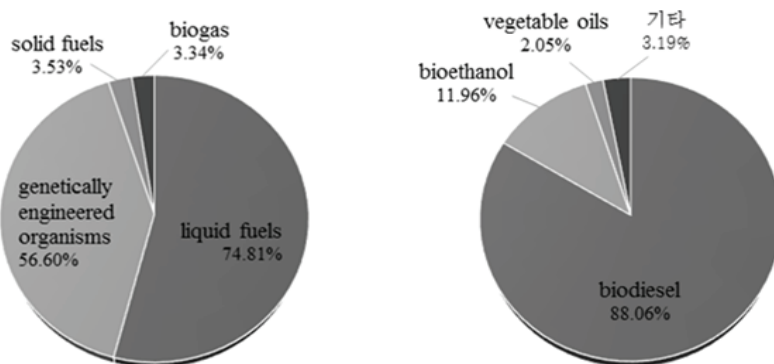
〈표 1〉 분야별 IPC

Bio-fuels	IPC
Solid fuels	C10L 5/00, C10L 5/40-5/48
•Torrefaction of biomass	C10B 53/02, C10L 5/40, 9/00
Liquid fuels	C10L 1/00, 1/02, 1/14
•Vegetable oils	C10L 1/02, 1/19
•Biodiesel	C07C 67/00, 69/00, C10G, C10L 1/02, 1/19, C11C 3/10, C12P 7/64
•Bioethanol	C10L 1/02, 1/182, C12N 9/24, C12P 7/06-7/14
Biogas	C02F 3/28, 11/04, C10L 3/00, C12M 1/107, C12P 5/02
From genetically engineered organisms	C12N 1/13, 1/15, 1/21, 5/10, 15/00, A01H

〈그림 4〉 연도별 등록 특허 수 및 해당 분야 등록 특허 수



〈그림 5〉 연료별 및 액체 연료의 특허 구성 비율



2. 세부 기술요소 정의

바이오 연료와 관련된 등록 특허를 수집한 후 바이오 연료 기술에 대한 세부구성 기술요소를 파악하기 위하여 본 연구에서는 토픽모델링 기법을 활용하였다. 최종 선별한 유효 특허 집합에서 총 21,274개의 바이오 연료의 기술적 의미를 포함하는 키워드 집합을 구성하였다. 그 후 빈도를 기반으로 하는 특허-키워드 매트릭스를 구축하였고 이는 토픽모델링에서 바이오 연료의 세부 구성 기술을 파악하기 위한 투입자료로 활용된다. 본 연구에서 확률적 모형인 토픽모델링 알고리즘 중 성능이 가장 우수한 것으로 알려진 LDA모형을 활용하였으며, LDA모형을 실시하기에 앞서 전체 특허 내 세부 기술을 나타내는 토픽의 개수를 결정해야 한다. 최적 토픽 수는 모든 토픽 주제가 명확히 구분되도록 결정하기 위하여 토픽 간의 연결성을 기반으로 토픽 간 유사도가 가장 낮아야 한다. 이를 위해 본 연구에서는 코사인 유사도(Cosine Similarity)를 사용하였다. 코사인 유사도는 두 벡터 간의 유사도를 산출하는 가장 널리 알려진 방법이며 벡터 A와 벡터 B의 코사인 유사도의 공식은 다음의 식 2와 같다.

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2 \times \sum_{i=1}^n (B_i)^2}} \quad (\text{식 2})$$

본 연구에서는 키워드들을 상호관련이 존재하는 기술 주제로 가장 잘 분리하는 최적 토픽 수를 결정하기 위한 방법으로 LDA모형의 결과물 중 하나인 토픽-키워드 매트릭스를 활용하였다. 토픽-키워드 매트릭스는 각 토픽 주제를 키워드 확률 벡터로 표현한 것이며, 키워드가 토픽에 기여하는 정도를 파악하여 각 토픽의 주제를 결정하는 데 사용된다. 이때 토픽-키워드 매트릭스의 전체 토픽 간 평균 유사도가 가장 낮은 토픽 수를 선택할 수 있다. 최적 토픽 수를 결정하기 위하여 토픽 수를 10부터 60까지 설정한 후 토픽모델링을 실시하였다. 그 후 토픽 간 평균 코사인 유사도를 산출한 결과 평균 코사인 유사도가 0.009814로 가장 낮은 수치를 가지는 토픽 수는 49개임을 확인할 수 있다. 최적 토픽 수 49개에 해당하는 토픽-키워드 매트릭스의 확률 분포를 활용하여 표 2와 같이 각 토픽을 구성하는 키워드 가운데 높은 확률 값을 가지는 상위 키워드를 파악하고 해당 토픽을 포함하는 특허문서를 참고하여 각 토픽이 나타내는 주제를 정의하였다. 이때 토픽의 주제는 연구 분야인 바이오 연료와 관련된 세부 구성 기술을 의미하게 된다. 이러한 레이블링 과정을 거쳐 정의된 토픽들은 다음 과정인 네트워크 분석과 정보 엔트로피 분석을 진행하는 데 활용된다.

〈표 2〉 토픽별 상위 키워드와 레이블링 결과

Label	Topic	1 st 키워드	1 st Prob	2 nd 키워드	2 nd Prob	3 rd 키워드	3 rd Prob	4 th 키워드	4 th Prob
아미노산 구조결정	40	amino acid residue	0.152	nucleic acid molecules	0.15	amino acid residues	0.142	amino acid sequences	0.075
폐수처리	41	reaction vessel	0.027	feed stream	0.025	effluent stream	0.021	hydrogen sulfide	0.02
중유	42	heavy oil	0.121	fluorescent protein	0.051	slurry catalyst	0.042	catalyst precursor	0.034
젖산생성 유전자	43	DNA segment	0.061	lactic acid	0.038	tomato plant	0.034	nucleic acid segment	0.026

3. 세부 기술 간 네트워크 생성 및 분석

본 단계는 네트워크 분석을 활용하여 총 34,237개의 특허 문서를 4개의 연도구간으로 나누어 바이오 연료 관련 세부 기술의 연관관계를 시각화하여 핵심 기술 및 기술 간 상호관계 파악을 진행한다. 네트워크 분석을 진행하기 위하여, 특허-토픽 매트릭스에서 특허가 등장한 시점인 출원일 기준으로 연도구간을 나누어 바이오 연료 관련 기술 동향을 파악하고자 한다. 총 4개의 연도구간 A(2000년 이전), B(2001~2005), C(2006~2010), D(2011~2015)로 나누어 진행하였으며 각 연도구간별로 최종 특허-토픽 매트릭스를 생성한다. 그 후 각각의 특허-토픽 매트릭스를 활용하여 토픽 간 동시 출현을 나타내는 총 4개의 토픽-토픽 동시출현 매트릭스를 생성한다. 동시출현 분석(Co-occurrence analysis)이란 기술문서의 키워드가 해당 기술문서의 내용을 적절하게 표현하고 있다는 가정하에 이루어지며, 따라서 같은 기술문서 내에서 한 쌍의 키워드가 동시에 발생한다는 것은 이들이 서로 연관되어 있음을 의미한다.¹⁴⁾ 또한 같은 쌍의 키워드들이 다양한 기술문서 내에서 함께 자주 나타난다는 것은 해당 키워드가 해당 기술 분야에서 밀접한 관련성을 지니고 있다는 것을 말한다.¹⁵⁾ 다음 표 3과 같이 생성된 최종 토픽-토픽 동시출현 매트릭스는 해당된 두 토픽 간 동시출현하는 연결강도를 나타낸다.

표 3과 같이 생성된 토픽-토픽 동시출현 매트릭스는 네트워크 형태의 시각적 표현이 가능하며 연도구간별로 토픽의 동향 분석이 가능하게 된다. 토픽 네트워크의 노드는 각 토픽, 즉 바이오 연료 관련 세부 주제이며 링크는 세부 주제별 연결 관계를 나타낸다. 각 노드의 크기는 해당 토픽을 포함하는 바이오 연료 관련 등록 특허 수를 의미하며, 링크의 굵기는 해당 토픽 간 연

14) Cambrosio, Alberto et al., "Historical Scientometrics? Mapping over 70 Years of Biological Safety Research with Coword Analysis", *Scientometrics*, Vol.27 No.2(1993), pp.119-143.

15) 서원철 외 2인, "단오동시출현분석을 통한 한국의 국가 R&D 연구동향에 관한 탐색적 연구", 『Journal of Information Technology Applications & Management』, 제19권 제4호(2012), 1-18면.

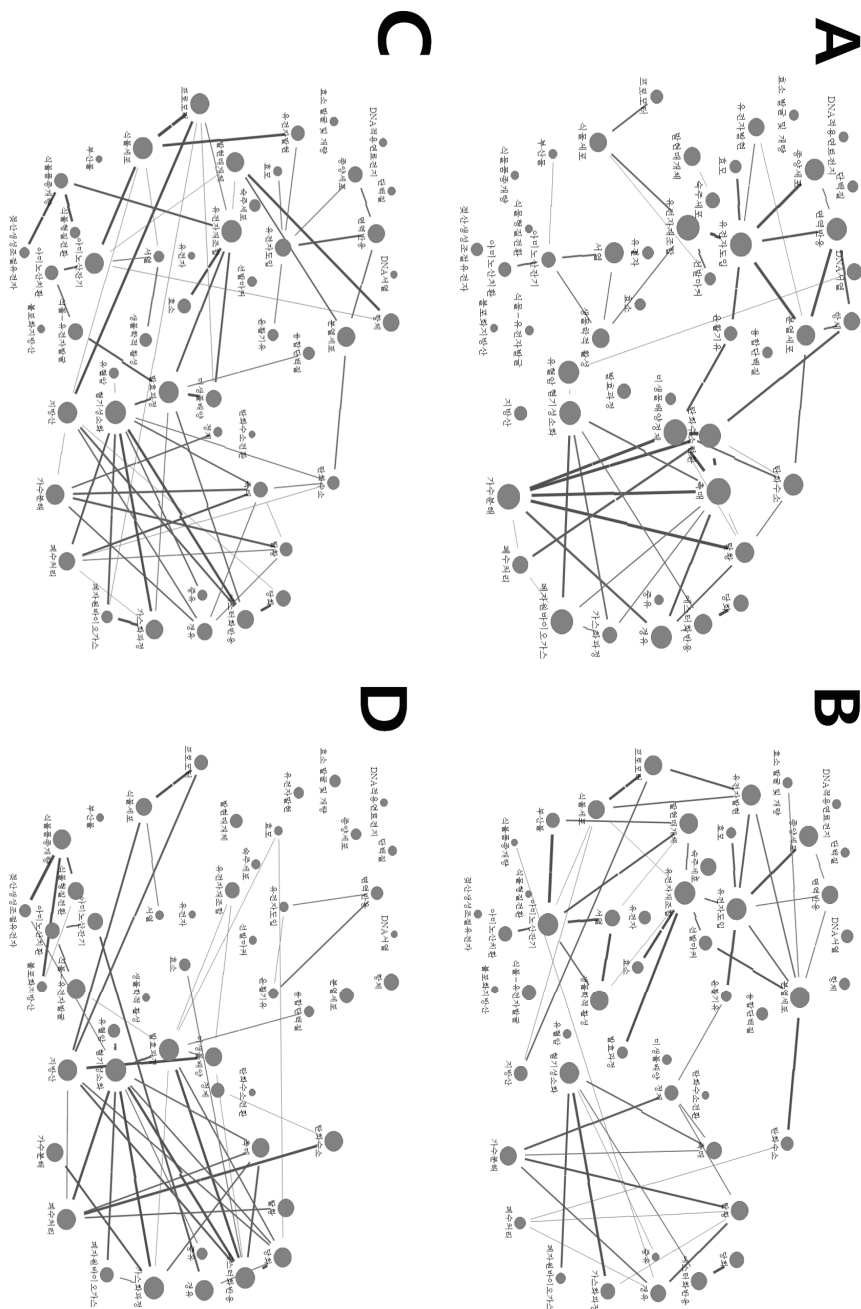
〈표 3〉 토픽-토픽 동시출현 매트릭스

	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10
Topic1	0	0,8199	0,3890	0,0442	0,0729	0	0	0	0	0,6720
Topic2	0,8199	0	0	0,1318	0,2693	0	0	0,0945	0	0,0356
Topic3	0,3890	0	0	0,1264	0,1206	0,2001	0,1353	0	0	0,1998
Topic4	0,0442	0,1318	0,1264	0	0,1560	0,2333	0	0,0389	0	0,0630
Topic5	0,0727	0,2693	0,1206	0,1560	0	0,1396	0,2364	0,1628	0	0

결 관계의 강도를 나타낸다. 전체 세부 노드 간 연결 관계의 복잡도가 높아 시각적 이해도가 낮아지는 것을 방지하기 위하여 상위 10%의 높은 연관관계를 가지는 링크만 산출하여 효율적으로 시각화 분석을 진행하였다.

4개의 연도구간의 토픽-토픽 동시출현 매트릭스를 활용하여 바이오 에너지 분야의 세부 기술 간 연결 관계 네트워크를 그림 6과 같이 생성할 수 있다. 전반적인 바이오 연료 관련 세부 기술의 네트워크 분석과 각 구간별 토픽을 포함하는 특허 수와 토픽 간 연결 관계의 강도 등 계량적 수치를 통한 분석을 진행하였다. 연도구간 A에는 ‘축매’, ‘정제’, ‘유전자도입’, ‘가수분해’ 등과 같은 토픽들이 특허 수가 1297건, 988건, 900건, 799건으로 상위를 차지하였고, ‘정제’와 ‘탄화수소 전환’의 링크 값은 6.903으로 가장 높은 연관관계를 가지며 이는 두 토픽을 동시에 포함하는 특허의 수가 많다는 것을 의미한다. 이와 같이 2000년대 이전에는 바이오 연료 관련 원천 기술의 개발에 연구가 집중되어 발전하는 모습을 확인할 수 있다. 이후 2000년대에 진입하면서 연도구간 B에서는 ‘유전자 재조합’을 포함하는 특허 수가 452건으로 가장 높은 수를 차지하였으며 ‘유전자 도입’, ‘발현매개체’ 등의 세부 기술과 관련된 새로운 연결 관계가 생성되며, 연료 생성의 효율성을 위한 유기체 유전 조작에 관한 유전공학 관련 토픽의 특허 수 또한 증가하는 경향을 보이고 있다. 연도구간 C 이후 ‘혐기성소화’를 포함하는 특허 수가 487건으로 바이오 연료 관련 핵심 기술로 떠올랐으며, ‘유전자재조합’의 특허 수 또한 365건으로 높은

〈그림 6〉 연도구간별 바이오 연료 세부기술 네트워크



것으로 보아 유전공학 분야의 기술 개발과 미세조류와 관련된 기술이 함께 발전하고 있음을 알 수 있다. D구간에서는 361건으로 ‘혐기성소화’의 뒤를 이어 ‘가스화’, ‘발효과정’을 포함하는 특허 수가 277건, 272건으로 높은 비중을 차지하고 있는 것으로 보아 최근 급부상하고 있는 미세조류와 관련된 연구가 중점적으로 진행되고 있음을 판단할 수 있다. 이러한 네트워크의 변화는 최근 고부가가치 산업을 위하여 식물성 플랑크톤으로 불리는 미세조류와 관련된 연구가 급부상하고 있는 점을 반영하고 있다. 미세조류는 석유 고갈 시 대체자원 및 신재생 에너지로 가장 효율이 높다는 연구가 진행되며 전 세계적으로 경제성, 생산성을 높이기 위한 개발이 진행 중이다.¹⁶⁾¹⁷⁾ 따라서 앞으로 미세조류를 활용한 바이오 연료 생산의 경제성을 확보하기 위하여 여러 나라에서 원천 기술과 특허 확보에 노력을 집중할 것으로 예상된다.

4. 기술요소별 정보 엔트로피 분석

본 단계에서는 각 세부 기술이 다른 기술 분야와의 다양성 및 활용성을 측정하고 향후 발전가능성을 분석하기 위하여 기술요소별 정보 엔트로피를 활용하였다. 네트워크 분석을 통하여 직관적인 이해와 전체적인 동향을 살펴 보았으며, 정보 엔트로피의 분석을 통하여 계량적 지표를 활용한 각 세부 기술의 연결 관계에 관하여 심도 있게 분석하였다. 연도구간별 상위 엔트로피 값을 가지는 세부 기술의 변화와 해당 주요 토픽이 다른 토픽과 어떤 연관을 맺고 있는지 기술적인 의미를 파악하였다. 각 기술의 엔트로피를 측정하기 위하여 연도구간별 토픽-토픽 동시출현 매트릭스를 활용하였으며 연도구간별로 서로 비교 분석이 가능하도록 전체 연도구간에 대한 전체 링크의 합을 각 연도구간의 개별 링크 값에 나누어 확률 값으로 표현하였다. 그 후 식 1을

16) Hannon, Michael et al., “Biofuel from algae: challenges and potential”, *Biofuels*, Vol.1 No.5(2010), pp.763-784.

17) Milano, Jasssinnee et al., “Microalgae biofuels as an alternative to fossil fuel for power generation”, *Renewable and Sustainable Energy Reviews*, Vol.58(2016), pp. 180-197.

적용하여 전체 49개 토픽의 연도구간별 정보 엔트로피 값을 산출한 후 모든 토픽의 평균 절댓값과 증가율의 기하평균을 계산하여 표 4와 같이 생성하였다.

구간 A는 ‘유전자도입’, ‘축매’, ‘정제’의 정보 엔트로피가 각각 0.086, 0.084, 0.078의 상위 값을 가지는 것을 확인할 수 있으며, 각각 분야의 발전과 성장을 위한 원천기술의 연구가 진행되고 있으며 원천기술을 중심으로 다양한 기술 개발이 진행되고 있음을 판단할 수 있다. 구간 B에서는 ‘아미노산 잔기’, ‘유전자재조합’, ‘유전자도입’의 값이 0.080, 0.078, 0.058로 높은 엔트로피 값을 가지며 유전공학과 관련된 토픽이 상위에 차지하였다. 구간 C에 진입한 후 ‘혐기성소화’가 0.077로 상위 정보 엔트로피 값으로 등장하기 시작하였으며 ‘유전자재조합’, ‘식물세포’ 등 유전공학 분야도 0.071, 0.066의 값으로 꾸준히 발전하였다. 구간 D부터는 유전공학 영역의 기술적 연계성이 하락하였으며 미세조류를 활용한 바이오 연료 생산 개발이 진행되면서 그와 관련된 세부 기술들이 중점적으로 다양한 연구들과 결합되어 발전하였다. 이와 관련된 토픽들은 ‘혐기성소화’, ‘지방산’, ‘발효과정’으로 각각 0.128, 0.104, 0.098의 값으로 높은 정보 엔트로피를 가져 높은 활용성을 보이며 발전하였다.

상위수준의 정보 엔트로피 값을 가지는 바이오 연료 기술의 전반적인 분석 외에, 본 연구에서는 표 4의 해당 토픽별 정보 엔트로피의 전체 연도구간 평균값과 증가율을 x축, y축으로 정의하고 2차원상에 나타낸 후 각각의 평균값인 0.03603과 0.14462로 4분할하여 그림 7과 같이 엔트로피 맵을 구성하였다. 영역 A는 평균 엔트로피 값과 증가율의 기하평균 값이 모두 평균 이상의 값을 가진다. 해당 토픽이 현재 다른 기술 분야와 다양하게 연계되며 동시에 높은 활용성을 가져 향후 발전가능성이 높아 다른 기술 분야와 연계적인 측면에서 발전이 지속될 것을 의미한다. 영역 A에 해당하는 세부 기술로는 ‘혐기성소화’, ‘지방산’, ‘발효과정’ 등 9개의 기술 중 미세조류 토픽이 주로 포함되며, 이는 향후 미세조류와 관련된 기술이 지속적으로 발전할 것을 예상한다. 이러한 결과가 현재 논의되고 있는 바이오 연료 기술개발 전략에

부합하며 바이오 연료의 효율성과 생산성을 높이기 위하여 여러 나라와 기관에서 미세조류를 활용한 연구가 중점적으로 이루어질 것임을 의미한다.¹⁸⁾ 영역 B는 엔트로피의 평균값은 낮으나 증가율의 기하평균이 높아 발전가능성을 가진 기술을 의미하며, ‘식물품종개량’, ‘불포화지방산’, ‘식물형질전환’을 포함하여 총 11개의 토픽들이 위치한다. 이 영역은 성장성이 높으나 영역 A로 발전하기 위해서는 높은 수준의 연구와 개발이 요구되며 투자비용이 발생할 수 있다. 해당되는 토픽들 중 평균선과 가까이 위치하지 않는 기술 분야는 식물과 연관되어 있는 주제이며, 2000년대 이후 유전공학 분야 기술의 성장으로 인하여 식물을 유전적으로 조작하는 형질전환과 교배를 통한 품종개량 기술의 엔트로피 증가율이 상승한 것으로 예상되며 영역 A로의 이동을 위한 적극적인 투자와 연구를 진행하거나 연구 개발을 중단하여 핵심 기술의 발전에 집중할 것인지에 관한 전략 생성이 필요하다. 영역 C는 평균 엔트로피 값과 증가율 모두 평균 이하의 값을 가지며, 15개의 세부기술이 위치한다. 이 영역은 현재 핵심기술에 해당하지 않고 더 이상의 발전을 기대할 수 없는 기술로 효율적인 바이오 연료 생산을 위하여 다른 영역 기술의 발전에 관심을 가져야 한다. 마지막으로, 영역 D는 연도구간마다 높은 정보 엔트로피 값을 가지지만 높은 성장률은 보이지 않는 기술 분야가 위치하며, 이미 바이오 연료 생산에 있어 다른 기술들과 높은 연계성을 가지는 주요 분야라고 할 수 있다. 14개의 기술들이 포함되며 ‘유전자재조합’, ‘유전자도입’, ‘축매’, ‘정제’와 같은 기술들이 위치한다. 유전자 관련 세부 기술들은 바이오 연료 생산에 있어 제외할 수 없는 핵심 기술이므로 다양한 기술들과 연관되어 있으며 한정된 바이오매스에서 효율적으로 연료를 생산하기 위해서는 유전공학 기술의 응용이 필요하다. 또한 화학반응을 돕는 촉매와 불순물 제거를 위한 정제 기술은 화학, 생물공학에서 필수적으로 요구되는 기술로 바이오 연료 제조 과정의 다양한 기술에 활용되고 있다. 이 영역을 활용하여 영역 A

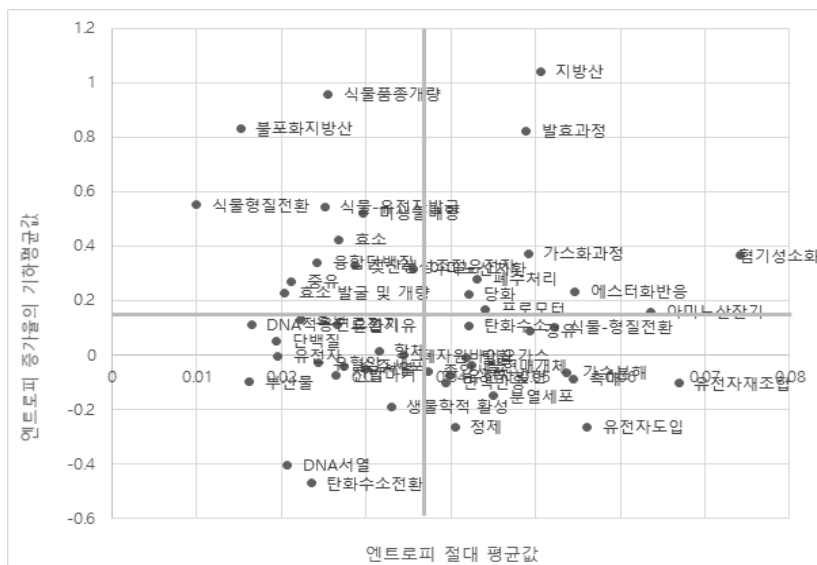
18) Subhadra, Bobban & Edwards, Mark, “An integrated renewable energy park approach for algal biofuel production in United States”, *Energy Policy*, Vol.38 No.9 (2010), pp.4897-4902.

와 영역 B에 해당하는 세부 기술의 발전을 이끌 수 있으며 향후 바이오 연료 산업의 성장에 대비할 수 있다.

〈표 4〉 토픽의 연도구간별 정보 엔트로피 값

Topic	A	B	C	D	평균값	증가율	Topic	A	B	C	D	평균값	증가율
1	0.014	0.014	0.040	0.051	0.030	0.518	26	0.020	0.017	0.023	0.029	0.022	0.130
2	0.012	0.027	0.059	0.104	0.051	1.042	27	0.041	0.047	0.040	0.032	0.040	-0.076
3	0.031	0.025	0.027	0.027	0.027	-0.043	28	0.013	0.014	0.018	0.020	0.016	0.111
4	0.032	0.034	0.027	0.027	0.030	-0.052	29	0.048	0.034	0.041	0.035	0.039	-0.100
5	0.019	0.016	0.020	0.023	0.019	0.053	30	0.005	0.001	0.014	0.019	0.010	0.550
6	0.045	0.036	0.031	0.037	0.037	-0.059	31	0.042	0.039	0.029	0.022	0.033	-0.191
7	0.042	0.031	0.038	0.057	0.042	0.107	32	0.054	0.030	0.042	0.070	0.049	0.090
8	0.032	0.013	0.023	0.029	0.024	-0.030	33	0.054	0.050	0.043	0.033	0.045	-0.151
9	0.006	0.008	0.031	0.056	0.025	0.959	34	0.041	0.022	0.011	0.009	0.021	-0.405
10	0.022	0.016	0.018	0.022	0.020	-0.006	35	0.078	0.027	0.026	0.031	0.040	-0.264
11	0.068	0.078	0.072	0.049	0.067	-0.105	36	0.015	0.023	0.059	0.099	0.049	0.821
12	0.016	0.032	0.006	0.011	0.016	-0.100	37	0.042	0.021	0.034	0.041	0.034	-0.002
13	0.014	0.015	0.023	0.030	0.020	0.227	38	0.010	0.020	0.030	0.040	0.025	0.544
14	0.014	0.019	0.028	0.035	0.024	0.339	39	0.027	0.043	0.064	0.042	0.044	0.166
15	0.023	0.028	0.038	0.053	0.036	0.315	40	0.044	0.080	0.062	0.068	0.064	0.158
16	0.051	0.034	0.032	0.049	0.042	-0.010	41	0.036	0.019	0.043	0.075	0.043	0.281
17	0.038	0.026	0.036	0.069	0.042	0.223	42	0.016	0.013	0.024	0.032	0.021	0.270
18	0.033	0.027	0.050	0.087	0.049	0.373	43	0.021	0.014	0.031	0.048	0.029	0.328
19	0.044	0.040	0.067	0.059	0.052	0.104	44	0.049	0.030	0.048	0.091	0.055	0.234
20	0.005	0.009	0.016	0.031	0.015	0.831	45	0.085	0.027	0.042	0.064	0.054	-0.087
21	0.011	0.026	0.038	0.032	0.027	0.422	46	0.072	0.008	0.003	0.011	0.024	-0.471
22	0.027	0.027	0.030	0.022	0.026	-0.076	47	0.041	0.042	0.050	0.037	0.042	-0.039
23	0.086	0.059	0.045	0.034	0.056	-0.265	48	0.027	0.020	0.022	0.037	0.027	0.109
24	0.072	0.037	0.047	0.058	0.054	-0.066	49	0.050	0.041	0.078	0.128	0.074	0.365
25	0.036	0.021	0.032	0.037	0.031	0.014							

〈그림 7〉 엔트로피 맵



V. 결론 및 추후연구

심각한 환경오염과 화석연료 고갈 문제로 인하여 신재생 에너지가 주목받으며 친환경성과 다양한 활용성을 지닌 바이오 에너지에 대한 관심이 집중되었다. 바이오 에너지는 자연의 유기 생물체를 활용하여 변환한 연료로 유전공학, 합성생물학, 화학공학 등 다양한 분야의 학문이 결합되어 활발한 연구 개발이 이루어지고 있다. 본 연구에서는 기술에 대한 상세한 자료가 포함되어 있는 특허 문서를 통한 객관적인 분석을 진행하였다. 다량의 특허를 수집하여 해당 문서의 전문 내용을 기반으로 바이오 연료 분야를 구성하고 있는 세부 기술요소를 파악하고, 연도구간별 기술요소의 동향을 살펴보았다.

본 연구에서는 바이오 연료 기술 관련 IPC 범위 내 해당하는 USPTO에 등록된 34,237건의 특허 문서를 수집하였으며 특허 전문으로부터 기술적 의미를 지니는 키워드를 추출하였다. 추출된 키워드를 활용하여 토픽모델링을

진행하였고 총 49개의 세부 기술요소를 정의하였다. 바이오 연료 관련 기술의 동적인 변화를 살펴보기 위하여 기술요소의 동시출현 매트릭스를 4개의 연도구간으로 나누어 구축한 후 네트워크 형태로 시각화한 후 네트워크를 살펴보며 각 구간마다 핵심적인 기술요소와의 연결 관계를 파악하였다. 시각화된 네트워크 분석은 단순한 데이터로는 파악할 수 없는 전체적인 특징과 연결 관계의 변화를 한눈에 확인할 수 있다. 마지막으로, 각각의 연도구간마다 세부 기술요소의 정보 엔트로피를 파악하여 상위 값을 가지는 기술분야의 변화를 살펴본 후 엔트로피 맵을 구축하였으며 각 영역마다 기술의 특징을 살펴보았다.

본 연구에서 분석한 결과는 바이오 연료를 생산하는 데 있어 관련된 세부 기술요소를 파악할 수 있으며 기술의 전체적인 동향 분석 결과를 제공한다. 이는 바이오 연료를 연구하는 다양한 분야의 연구자들과 기업에게 기초 연구 자료로써 활용될 수 있으며 향후 연구 개발의 방향성을 설정하는 데 효과적일 것으로 기대된다. 하지만 본 연구는 몇 가지 현실적인 한계점과 그에 따른 향후 연구의 과제를 지닌다. 우선 특허 문서에서 키워드를 추출한 후 전처리 과정을 통하여 키워드 집합을 구축할 때 불용어 제거 기준과 최적 토픽 수의 선정 기준에 따라 세부 기술요소의 정의가 변할 수 있다. 다음으로 분석의 신뢰성 측면에서 등록 특허 문서의 출원일을 기준으로 분석을 진행하였기 때문에 최신의 특허 데이터는 동향 분석에 통합되지 않아 최신 연구 동향을 분석하기에는 한계점을 지닌다. 또한 특허 내 유효 토픽만을 선별하기 위한 특허-토픽 매트릭스의 적절한 임계값을 선정하는 방법에 대한 후속연구가 필요할 것이다. 마지막으로 본 연구는 기존 개발된 방법론을 바이오 연료 분야에 적용하여 분석하였으나 향후 연구에서는 기술 모니터링에 있어 방법론 개선, 새로운 이론 제시를 수행하여 분석할 예정이다.

참고문헌

〈단행본(국내 및 동양)〉

김용학, 『사회연결망분석』, 박영사, 2003.

〈단행본(서양)〉

Birol & Fatih, *World Energy Outlook*, Paris: International Energy Agency, 2008.

〈학술지(국내 및 동양)〉

김갑수 외 2인, “엔트로피를 이용한 객체 지향 프로그램의 복잡도 척도”, 『정보 학회는
문지(B)』, 제22권 제12호(1995).

민은주 · 김수덕, “수송용 바이오에탄올 도입의 경제성 평가 및 CDM 사업 타당성 분
석”, 『에너지경제연구』, 제7권 제1호(2008).

서원철 외 2인, “단오동시출현분석을 통한 한국의 국가 R&D 연구동향에 관한 탐색적
연구”, 『Journal of Information Technology Applications & Management』, 제19
권 제4호(2012).

송효학 · 박종명, “바이오에너지 산업의 현황 및 최신 연구동향”, 『BT NEWS』, 제20권
제2호(2013).

이정, “정보이론의 엔트로피 관점에서의 바라본 온라인 소비자 리뷰의 소비자 의사결
정에 있어 불확실성 감소 효과”, 『한국전자거래학회지』, 제16권 제4호(2011).

정병기 외 2인, “융합기술의 동향분석을 위한 의미론적 특허분석 접근 방법”, 『지식재
산연구』, 제11권 제4호(2016).

최일영 · 이현수, “[기업경영연구] 에서의 공동연구 네트워크 분석”, 『기업경영연구』,
제56권 단일호(2014).

〈학술지(서양)〉

Blei et al., “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*,
Vol.3 No.Jan(2003).

Cambrosio et al., “Historical Scientometrics? Mapping over 70 Years of Biological
Safety Research with Coword Analysis”, *Scientometrics*, Vol.27 No.2(1993).

Jeong Byeongki & Yoon Janghyeok, “Competitive Intelligence Analysis of
Augmented Reality Technology Using Patent Information”, *Sustainability*, Vol.9
No.4(2017).

- Yoon Byungun & Park Yongtae, "A Text-Mining-Based Patent Network: Analytical Tool for High-Technology Trend", *The Journal of High Technology Management Research*, Vol.15 No.1(2004).
- Yoon Janghyeok et al., "Identifying Product Opportunities Using Collaborative Filtering-Based Patent Analysis", *Computers & Industrial Engineering*, Vol.107(2017).
- Schenk et al., "Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production", *Bioenergy Research*, Vol.1 No.1(2008).
- Hannon, Michael et al., "Biofuel from algae: challenges and potential", *Biofuels*, Vol.1 No.5(2010).
- Milano, Jasssinnee et al., "Microalgae biofuels as an alternative to fossil fuel for power generation", *Renewable and Sustainable Energy Reviews*, Vol.58(2016).
- Subhadra, Bobban & Edwards, Mark, "An integrated renewable energy park approach for algal biofuel production in United States", *Energy Policy*, Vol.38 No.9(2010).

〈연구보고서〉

안두현 외 3인, "바이오에너지 연구개발 동향과 시사점", 과학기술정책연구원, 2007.

Monitoring Bio-fuel Technology Using Patent Text Mining

Lee Dahye, Choi Hayoung, Jeong Byeongki & Yoon Janghyeok

Bio-fuel is a renewable energy which is attracting great attention owing to its environment-friendliness and usability, as environmental pollution makes efforts for renewable energy expand globally. Compared to the great interest in bio-fuel technology, however, few or no quantified analyses to identify this technology's sub-technologies and monitor their evolving trends have been conducted by previous studies. Therefore, this study analyzes evolving trends and characteristics of bio-fuel technology's sub-technologies based on patent data. The analysis process of this study is organized as follows: 1) patents related to bio-fuel technology are collected and keywords are then extracted from the text of the patents, 2) Topic modeling technique with the keywords is applied to patent documents to identify sub-technologies constituting bio-fuel technology, 3) trends by topic co-occurrence networks with respect to the sub-technologies are visualized and examined, and 4) information entropy analysis is applied to identify the connection strength, diversity and future promise of the sub-technologies. Through this study, evolving trends of bio-fuel technology can be identified and the interrelationship among its sub-technologies can be analyzed. The results of this study can be used for companies and researchers to understand the changing process in bio-fuel technology field and thereby to determine further research directions.

Keyword

Bio-fuel technology, Technology monitoring, Text mining, Topic modeling, Network analysis, Information entropy