

# 구글 검색과 특허정보를 활용한 지속가능성 평가 방법 연구

지선구

국립금오공과대학교 IT융합학과

## A Study on Sustainability Evaluation Method Using Google Search Results and Patent Information

Seonkoo CHEE

Department of IT Convergence, Kumoh National Institute of Technology

**요약** 본 연구는 특정 분야나 기업 등에 대한 지속가능성을 평가하기 위한 방법론의 하나로, 트리플 버텀 라인(TBL: Triple Bottom Line) 비즈니스 프레임 워크의 각 축 별 지속가능성을 자동으로 평가할 수 있는 새로운 접근 방식을 제시하는 것을 목적으로 한다. 본 연구의 평가 방법은 먼저, 텍스트 마이닝을 통해 지속가능성 사전에 구축하며, 사전과 평가 대상으로부터 추출한 단어 집합과의 코사인 유사도를 산정하여 지속가능성을 계산한다. 이에 더해, 구글 검색 데이터를 이용해 사전에 포함된 단어 각각의 관계성을 고려한다. 본 연구에서 제시하는 지속가능성 평가 방법을 검증하기 위해 지속가능성 경향성이 알려진 학술논문 및 뉴스 기사를 대상으로 테스트한 결과, 본 방법은 TBL의 각 축에 대해 지속가능성 경향을 정량적으로 합리적으로 예측하는 것으로 나타났다. 또한, FDM(fused deposition modeling) 방식의 3D 프린팅에 대해 출원된 특허 데이터에 본 평가 방법을 적용한 결과, 3D 프린팅 기술 자체뿐만 아니라 그 3D 프린팅 기술을 개발한 회사의 지속가능성을 평가하는 과정에도 유의미한 결과를 얻을 수 있는 것으로 나타났다. 본 연구의 평가 방법은 기업이나 정부기관 등이 TBL의 각 축 별 지속가능성을 정량적으로 평가하고 주기적으로 모니터링 하면서 지속가능성을 높이는 활동을 지속하도록 촉진하는 일종의 도구로 활용될 것으로 기대된다.

**Abstract** This study presents a new approach to sustainability assessment that can be used to automatically evaluate sustainability in each axis of the Triple Bottom Line (TBL) business framework for a specific field. A sustainability dictionary is built through text mining, and cosine similarity between the dictionary and a word set extracted from a target field is identified to determine sustainability. In addition, a relation factor for the words in the dictionary is considered by using Google search data. Evaluations of academic papers and news articles with a known sustainability tendency found that the proposed method reasonably predicted, in a quantitative manner, the tendency of sustainability in each TBL axis. When the proposed method was applied to patents for 3D printing, it was found that this approach also gives a plausible result when assessing the sustainability not just of 3D printing itself but of a company that develops 3D printing technologies. It is expected that the present approach will allow organizations to quantitatively evaluate their sustainability on each axis and enhance it.

**Keywords** : Sustainability, Triple Bottom Line, Text Mining, Cosine Similarity, Google Search Data

이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구지원사업의 지원을 받아 수행된 연구임(NRF-2020S1A5A8044514).

\*Corresponding Author : Seonkoo CHEE(Kumoh National Institute of Technology)

email: [lingerchee@kumoh.ac.kr](mailto:lingerchee@kumoh.ac.kr)

Received April 14, 2022

Revised May 11, 2022

Accepted July 7, 2022

Published July 31, 2022

## 1. 서론

환경 문제가 글로벌 이슈로 대두되면서 사회적 책임 측면에서 기업이 환경보호 등 우리 사회의 지속가능성(sustainability)에 대해 어떤 노력을 견지하고 있는지 관심이 집중되고 있다. 지속가능성이란 안녕(well-being)한 상태가 긴 기간 동안, 더 나아가 영원히 유지되는 것을 말한다[1]. 지속가능성이란 용어가 공식적으로 처음 도입된 것은 1987년 환경 관련 유엔위원회(UN World Commission on Environmental and Development)에서 작성한 보고서이다[2]. 초기에는 제한된 자원과 환경 파괴에도 불구하고 전 인류가 어떻게 좀 더 나은 생활을 유지해 나갈 것인가에 초점이 맞춰졌으나, 사회적, 경제적 측면까지 관심 영역이 넓어지면서 트리플 버텀 라인(TBL: Triple Bottom Line, 이하 TBL)이라는 비즈니스 원칙으로 발전하였다[1-3].

지속가능성이 중요해짐에 따라 국제기구인 GRI(Global Reporting Initiative)에서 지속가능성 보고 스탠다드(sustainability reporting standard)를 제정해 기업이나 정부기관 등이 지속가능성에 대한 현황을 공개하도록 권고하고 있다[4]. 이에 따라, 기업 등은 매년, 에너지 사용, 생물학적 다양성 유지, 노동 환경 개선 등 환경적(environmental), 사회적(social) 및 경제적(economic) 지속가능성에 관한 노력에 대해 지속가능성 보고서(CSR: Corporate Sustainability Report, 이하 CSR)를 만들어 배포하고 있다[5].

기업이나 정부기관의 활동이 어느 정도로 환경이나 사회를 양질로 유지하는지에 대해 평가하는 것은 상당히 중요한 이슈인바, 지속가능성 노력을 객관적이고 체계적으로 분석할 수 있는 평가 방법이 필요하다. 지속가능성 노력을 평가하는 방안 중 하나로 텍스트 마이닝(text mining) 기법을 이용해 지속가능성 관련 활동 정보를 정량적으로 분석하는 방법이 제시되었다[6-9]. 평가 객체가 되는 기업활동 정보로는 앞서 언급한 CSR이 될 수 있으며, 그 외 기업의 홈페이지, 홍보용 자료 등도 포함된다. 그러나 이런 자료들은 기업이 자사를 홍보하기 위해 가공하고 완화된 자료들로서 정보 왜곡이 있을 수 있어 지속가능성을 평가하기 위한 양질의 자료로 보기는 쉽지 않다. 그에 따라, 평가 결과도 부정확할 수 있는바, 기업 활동을 가감 없이 날 것(raw)의 상태로 알 수 있는 자료가 필요하다.

최근, 특허정보를 분석해 기업의 경영 전략을 수립하는 방법이 제안[10]된바, 경영 이론과는 다소 동떨어진

기술 자료를 경영 분야에 활용했다는 측면에서 주목받고 있다. 특히, 주목할 점은 특허정보는 기업이 개발한 기술을 특허권으로 유형화하기 위해 상세하게 서술한 문서로서 가공되지 않은 객관적인 자료라는 점이다. 본 연구에서는 기업의 지속가능성에 대한 평가 대상으로 특허정보를 활용하고자 한다. 비정형 문서 분석 방법인 텍스트 마이닝 기법으로 기업이 출원한 특허정보를 분석해 지속가능성 노력 수준과 TBL 지속가능성 분포를 확인할 수 있는 일련의 특허정보 기반 평가 방법론을 제안하고자 한다. 이때, 지속가능성 수준 평가를 위한 지표가 필요한데, 기존의 연구들과는 달리 구글 검색 결과를 텍스트 마이닝 기법에 연계시켜 지속가능성 평가 지표로 활용할 수 있는지 그 가능성도 타진하고자 한다.

## 2. 선행연구

### 2.1 지속가능성 평가

지속가능성 평가 방법으로 수동 콘텐츠 분석(MCA: Manual Content Analysis, 이하 MCA) 기법이 제안되었다[11, 12]. MCA는 CSR에서 컴퓨터 코드를 이용하거나 수동으로 지속가능성에 관한 항목 데이터를 추출한 후, 전문가가 다시 수작업으로 데이터를 검토해 지속가능성 정도(degree)를 확인하는 방법이다. 그러나 CSR의 크기가 급속도로 커지고 보고서를 발간하는 기업이 많아지면서 MCA로 분석하는 것은 비효율적이고 시간이 많이 소요된다는 문제가 있다. 또한, MCA는 수동으로 데이터를 분석하는 것이라 인적 오류 가능성을 내포하는 한계를 보였다[13].

이런 문제를 해결하기 위해 자동 콘텐츠 분석(ACA: Automated Content Analysis, 이하 ACA)의 일종으로 텍스트 마이닝 방법이 제안되었다[7-9]. 텍스트 마이닝은 비정형(unstructured) 또는 준정형(semi-structured) 문서 정보로부터 텍스트를 추출해 패턴을 읽어내는 자동화된 프로세스를 말한다. MCA와는 달리 인적 오류 가능성을 줄여 신뢰성이 높으며 저비용으로 다량의 데이터를 한꺼번에 분석할 수 있는 장점이 있다.

ACA는 대부분 평가 결과를 정량화하기 위한 사전(dictionary)을 필요로 한다. 사전이란 특정 분야를 대표하는 단어, 단어의 어근(words stem), 어구(phrase) 등을 카테고리화 함께 포함하는 단어의 집합인데, 평가 대상 정보나 문서와의 비교를 통해 사전에 속한 단어가 몇 개 어느 정도의 빈도(frequency)로 포함되어 있는지에

따라 카테고리를 분류하거나 중요도를 평가하게 된다[8]. 예를 들어, 어떤 문서가 사전에서 중요도를 높게 책정한 단어를 많이, 높은 빈도로 포함하고 있다면 타 문서에 비해 중요도가 높다고 판단하는 방식이다.

Deng 등은 2015년도에 발행된 IT분야 38개 기업 CSR을 기초로 302단어, 11개 카테고리로 구성된 단어 사전을 구축한 후, 지속가능성 관련 뉴스 449개를 대상으로 통계 분할표(contingency table)를 작성하였다[8]. 또한, 대응분석(correspondence analysis)으로 카테고리리와 보고서 작성 연도 간 관계를 조망하였다[8]. Modapothala와 Issac은 10개의 GRI 지속가능성 보고 스탠다드[10]에서 카테고리별 단어를 텍스트 마이닝으로 추출해 사전을 구축한 후, 기업들의 CSR에 포함된 키워드와 비교해 보고서 작성의 효율성을 평가하였다[6].

한편, Manuel은 독일의 16개 지자체 행동계획(municipal action plan)을 텍스트 마이닝해 경제, 재정, 식량, 에너지 등 카테고리 별로 단어를 배치한 후 동시 출현 네트워크 분석을 통해 지속가능성 요소간 관계를 분석하기도 하였다[14]. 또한, Shahi 등은 텍스트 마이닝과 기계학습 방법을 연계해 CSR에 대한 기계학습 모델을 구축한 후, 자동으로 기업의 CSR 작성 수준을 점수화하는 프로세스를 제시하였다[15,16].

## 2.2 특허정보 분석

특허정보는 기업이 생산하는 제품이나 기술을 가장 잘 알 수 있는 중요한 문서로서, 기술뿐만 아니라 서지정보, 기술의 개발 목적, 기술이 해결하고자 하는 제품의 문제점, 해결 방안 등 기업의 실체를 알 수 있는 객관적인 연계정보들을 포함하는 고품질의 활동 정보이다[10]. 이런 인식 아래 특허문서를 분석해 기업의 기술개발 전략을 수립하는 연구들이 다수 행해지고 있다[17,18]. Grimaldi 등은 기업의 특허정보가 연구개발 전략을 수립하고 기업의 기술 포트폴리오를 구축하는 데 유용함을 입증하고, 기업의 특허 포트폴리오를 질적 및 양적으로 분석해 기술개발 전략을 수립하는 새로운 방법론을 제안하였다[17]. 또한, Ernst 등은 시뮬레이션 분석을 기초로 특허정보가 연구개발에 어떻게 유용하게 활용되는지 실증하였다[18]. 동일한 프로젝트에 대해 특허로 보호받는 경우와 그렇지 않은 경우를 비교해 특허권을 받은 프로젝트의 경우 좀 더 큰 이익을 기대할 수 있음을 확인하였다. 이런 연구들을 통해 특허정보는 기업의 활동을 나타내는 핵심 정보임을 알 수 있고 특히, 기술정보 외에 해당 기술을 개발한 기업의 서지 정보를 알 수 있어 매우 가치가

높은 문서 정보임을 확인할 수 있다.

## 3. 지속가능성 평가 방법

### 3.1 지속가능성 평가 프로세스

본 연구는 지속가능성 평가 방법으로, 텍스트 마이닝 기법으로 지속가능성 사전(dictionary)을 구축하고, 평가 대상에서 추출한 단어 집합과 상호 관련성을 계산해 지속가능성을 측정하는 방법을 제안한다. 사전과 평가 대상 단어 집합의 관련성은 코사인 유사도를 사용하며, Fig. 1과 같이 구글 검색 데이터를 활용해 사전에 포함된 단어 각각의 중요도를 고려한다.

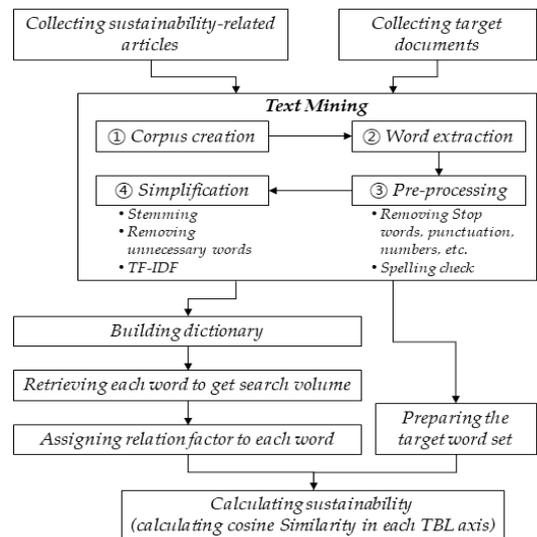


Fig. 1. Methodology

### 3.2 지속가능성 사전 구축

지속가능성 사전에 포함되는 단어들은 TBL의 3개 축(axis) 즉, 환경적, 사회적, 경제적 지속가능성을 대변하되 객관적으로 대표성을 인정받을 수 있어야 한다[8]. 본 연구는 지속가능성 분야의 대표 저널이면서 논문(article) 출간 수가 많아 모집단을 충분히 확보할 수 있는 「Sustainability」 저널을 시범적으로 주요 수집 대상으로 하였다. 「Sustainability」 저널을 검색할 수 있는 공개 시스템, MDPI[19]에서 지속가능성 각 축 별로 키워드 검색을 시행해 논문을 추출하였다.

다음으로, 논문들의 요약에서 텍스트 마이닝 기법으로 단어와 빈도를 구한다. 텍스트 마이닝은 R 데이터 언어

(R data language)를 사용하였으며, 코퍼스(corpus) 구축, 단어 추출, 전처리 및 간략화(simplification)를 위해 'tm', 'corpus', 'snowballc' 패키지(package)를 사용하였다[20]. 통상, 텍스트 마이닝 기법의 간략화 단계에서 Eq. (1)과 같은 TF-IDF를 이용해 불필요한 단어를 제거하거나 비중을 낮추는 바[8,21], 본 연구에서도 TF-IDF를 도입하기로 한다.

$$TF-IDF = \frac{freq(t_i, d_j)}{\sum_{k=1}^n freq(t_k, d_j)} \times \log\left(\frac{N}{count(t_i)}\right) \quad (1)$$

where,  $freq(t_i, d_j)$  denotes the frequency of the  $i$ -th word appearing in the  $j$ -th document,  $N$  is the total number of documents, and  $count(t_i)$  denotes the number of documents that contain the  $i$ -th word

### 3.3 사전 단어별 관계성 부여

본 연구는 구글 검색 데이터를 기초로 지속가능성 사전에 포함된 단어의 관계성을 부여한다. 어떤 두 개의 단어가 자주 같이 검색된다면 그 단어 간에는 중요한 연관 관계가 형성되어 있다고 가정하는 것이며, 동일한 문서에 공통으로 출현한 단어 간에는 밀접한 관계가 있다고 보고 문서를 분류하는 토픽 모델링(topic modeling)의 기본적인 가정과도 일치한다[22]. 예를 들어, 사전에 포함된 단어가 '탄소(carbon)'인 경우, 'environmental AND carbon'을 검색식으로 구글에서 검색하고 검색 건수 비율에 기반해 그 단어와 환경(environmental)과의 관계성을 얻을 수 있다.

Table 1. Exemplary calculation of relation factor using Google search volume

| Search expression       | environmental AND carbon | social AND carbon | economic AND carbon | Total |
|-------------------------|--------------------------|-------------------|---------------------|-------|
| Search volume (million) | 1,310                    | 1,250             | 441                 | 3,001 |
| Relation factor         | 0.44                     | 0.42              | 0.15                | 1     |

Table 1은 TBL 각 축 별로 탄소(carbon)라는 단어의 관계성을 예시한 것이다.

텍스트 마이닝을 통해 단어와 빈도를 추출하고 구글 검색 데이터를 이용해 관계성을 부여한 사전은 Table 2

와 같은 행렬(matrix) 형태를 갖는다. word #1의 경우 각 축에 따른 출현 빈도의 경향이 구글 검색 데이터를 기초로 한 관계성과 일치하나, word #2는 사회 축의 출현 빈도가 다른 축들과 유사하지만 사회축의 관계성은 매우 작은 값으로 계산되어 차이가 있음을 알 수 있다. 즉, word #2는 사전 구축에 사용된 사회적 지속가능성과 관련된 논문들에는 적지 않게 포함되어 있지만, 일반공중은 상대적으로 드물게 검색하는 단어라는 의미로 해석할 수 있다.

Table 2. Exemplary dictionary of the present study

| Words   | Frequencies |      |      | Relation factors |      |      |
|---------|-------------|------|------|------------------|------|------|
|         | Env.        | Soc. | Eco. | Env.             | Soc. | Eco. |
| word #1 | 0           | 0    | 27   | 0.07             | 0.07 | 0.86 |
| word #2 | 77          | 76   | 58   | 0.46             | 0.00 | 0.54 |
| word #3 | 96          | 113  | 52   | 0.19             | 0.44 | 0.37 |
| word #4 | 83          | 86   | 68   | 0.23             | 0.66 | 0.11 |
| ⋮       | ⋮           | ⋮    | ⋮    | ⋮                | ⋮    | ⋮    |
| word #n | 73          | 35   | 53   | 0.17             | 0.49 | 0.34 |

\*(Env.: Environmental, Soc.: Social, Eco.: Economic)

### 3.4 지속가능성 계산

본 연구의 사전(dictionary)은 Table 2에서 보는 바와 같이, 사전에 포함된 단어(word)별로 각 축에 대한 관계성(중요도)과 빈도를 원소(element)로 하는 행렬이다. 행렬의 각 열(row)은 지속가능성 축 별 벡터로 볼 수 있는데, 예를 들어, Table 2와 같은 사전에서 환경 축 지속가능성 벡터는  $R_{Env.}=(0.07, 0.46, 0.19, 0.23, \dots, 0.17)$  이 되며, 벡터의 길이  $n$ 은 사전에 포함된 단어의 총 개수이다.

지속가능성 평가 대상 문서들에 대해서도 Fig. 1의 프로세스를 따라 그 문서를 대표하는 단어들을 추출하면 평가 단어 벡터( $B$ )를 얻을 수 있으므로 미리 구축해 놓은 지속가능성 사전과 비교해 유사도를 구할 수 있다. 평가 단어 벡터  $B$ 는 특정 단어가 지속가능성 사전에 있는 경우 1, 없는 경우 0을 원소로 하는  $B=(1, 0, 0, 1, 1, \dots, 1)$ 와 같은 비트 벡터일 수도 있고, 또는 출현 빈도가 원소가 될 수도 있다. 평가 대상 데이터의 지속가능성은 평가 단어 벡터를 기초로 Eq. (2)와 같이 코사인 유사도를 이용해 계산한다[23].

$$Sustainability_k(R_k, B) = \frac{R_k \cdot B}{|R_k| \times |B|} \quad (2)$$

where,  $k=\{\text{environmental, social, economic}\}$

한편, 지속가능성 사전에 포함되는 단어 중 불필요한 단어의 비중을 낮추기 위해 각 단어 별 TF-IDF값을 기초로 Eq. (3)과 같이 지속가능성 벡터,  $R_k$ 를 수정(modify)한다[21].

$$R'_{k,i} = R_{k,i} \cdot (TF-IDF)_{k,i} \quad (3)$$

where,  $k$ ={environmental, social, economic},  $i$  = 1, ..., n and n is the total number of words in the dictionary

결국, 본 연구에서 지속가능성은 수정된 지속가능성 벡터를 이용해 Eq. (4)와 같이 계산한다.

$$Sustainability_k(R'_k, B) = \frac{R'_k \cdot B}{|R'_k| \times |B|} \quad (4)$$

where,  $k$ ={environmental, social, economic}

### 4. 지속가능성 평가 방법 검증

#### 4.1 경향성이 알려진 문서 대상 검증

본 연구의 지속가능성 평가 방법이 정상적으로 작동하는지 확인하기 위해 지속가능성 경향성이 알려진 문서 집합을 대상으로 지속가능성을 평가하였다. 검증을 위해 Fig. 1에 따라 구축한 지속가능성 사전의 주요 특징적인 단어들은 Table 3과 같다. MDPI(Molecular Diversity Preservation International)에서 추출[19]한 지속가능성 관련 논문 5,147개의 요약물 바탕으로 총 1,039개의 단어로 사전을 구축하였으며, 지속가능성 각 축 별로 900여 개의 단어가 고르게 포함되었음을 알 수 있다.

Table 3. Distinct words in each axis included in the dictionary

| Axis       | Environmental   | Social   | Economic   | Total |
|------------|---|--|--|-------|
| Words      | develop, produce, energy, manage, impact, urban, increase, technique, process, system, water, food, build, policy, improve, resource, industry, consume, etc. | manage, urban, produce, system, communicate, city, impact, policy, govern, energy, food, data, improve, local, design, strategy, project, etc. | produce, energy, manage, perform, urban, policy, system, increase, water, city, resource, grow, food, industry, data, agriculture, China, supply, farm, land, etc. | -     |
| # of words | 898   | 879  | 883  | 1,039 |

Table 4는 사전 구축의 기초가 된 각 축 별 논문 집합을 대상으로 지속가능성을 계산한 결과를 나타낸다. 환경 관련 지속가능성을 키워드로 검색한 논문에서 추출한 단어 집합은 환경 관련 지속가능성이 다른 축에 비해 현저하게 크게 계산되며, 사회 관련 지속가능성 논문의 단어 집합 및 경제 관련 지속가능성 논문의 단어 집합도 각각 사회 축, 경제 축의 지속가능성이 현저하게 높게 계산됨을 확인할 수 있다.

Table 4. Sustainability calculated using the present method: sustainability-related articles

| Test Document Set                                | Sustainability (% ratio) |                         |                         |
|--|--------------------------|-------------------------|-------------------------|
|  | Environmental            | Social                  | Economic                |
| Articles related to environmental sustainability | <b>0.2501</b><br>(71.1)  | 0.0357<br>(10.2)        | 0.0658<br>(18.7)        |
| Articles related to social sustainability        | 0.0186<br>(5.1)          | <b>0.2646</b><br>(72.9) | 0.0800<br>(22.0)        |
| Articles related to economic sustainability      | 0.0451<br>(11.7)         | 0.0712<br>(18.5)        | <b>0.2687</b><br>(69.8) |

다음으로, 특정 주제를 갖는 뉴스 기사를 대상으로 본 연구의 지속가능성 평가 방법을 검증하였다. 구글에서 뉴스 기사를 검색하되, 뉴스 기사들이 사회적 지속가능성 특성을 갖도록 'sustainability', 'social' 및 'health'를 검색어로 포함시키고 환경적, 경제적 이슈에 대한 기사는 제외되도록 검색식을 작성하였다. 여기서 사용한 구글 검색식은 'sustainability AND social AND health -finance\* -economy\* -environment\*'이며, 검색된 뉴스 기사를 기반으로 추출한 단어 집합에 대한 지속가능성 평가 결과는 Table 5와 같다.

Table 5. Sustainability calculated using the present method: news articles on health

| Test Document Set       | Sustainability(% ratio) |                         |                  |
|-------------------------|-------------------------|-------------------------|------------------|
|                         | Environmental           | Social                  | Economic         |
| News articles on health | 0.0649<br>(17.1)        | <b>0.1734</b><br>(45.7) | 0.1408<br>(37.1) |

Table 5에서 보는 바와 같이 사회적 지속가능성이 가장 크게 계산되었다. 구글에서 뉴스를 검색할 때 'social'과 'health'를 포함시키되 환경 및 경제 관련 기사는 제

외하도록 검색식을 만들었다는 사실과 환경적 지속가능성은 비교적 낮은 수준으로 계산되었다는 점을 고려하면 타당하게 예측한 것으로 볼 수 있다. 경제적 지속가능성도 사회적 지속가능성보다는 작으나 상당히 높게 계산되었는데, 건강에 대한 이슈가 경제와 밀접한 관계가 있기 때문으로 생각된다. 이상으로부터 본 연구에서 제시하는 지속가능성 평가 방법이 타당하게 작동하는 것을 확인할 수 있다.

#### 4.2 특허정보 대상 지속가능성 평가 방법 검증

일정한 기술적 특성을 갖는 특허정보를 대상으로 본 연구에서 제시하는 지속가능성 평가 방법을 검증하였다. 특허정보 검색 사이트[24]에서 특허분류(CPC: Cooperative Patent Classification, 이하 CPC)를 이용해 온실가스 및 관련된 특허 데이터만을 추출하였다. CPC는 유럽특허청이 제안한 특허분류코드의 일종으로, 국제특허분류(IPC: International Patent Classification, 이하 IPC)의 기술 분류를 조금 더 세분화한 것이다. CPC 중 'Y02C'는 '온실가스의 포집, 저장, 격리 또는 폐기하는 기술'에 대한 특허 데이터에 부여되는데, 총 1,017개의 특허 데이터가 검색되었으며 특허 데이터의 요약(abstract)만을 대상으로 검증 대상 단어 집합을 구성하였다. 또한, 다른 테스트 데이터 세트에 CPC, 'Y04S-10/58'가 부여되는 네트워크 기술의 회계적, 경제적 측면(financial or economic aspect related to network operation)에 대한 특허정보를 검색하였는데, 총 244개의 특허문서의 요약을 기초로 단어 집합을 구성하였다.

이렇게 만들어진 두 개의 검증 대상 단어 집합에 대한 본 연구의 지속가능성 평가 결과는 Table 6과 같다.

Table 6. Sustainability calculated using the present method: patent data selected using CPC

| CPC   | Sustainability (% ratio)       |                  |                                |
|---|--------------------------------|------------------|--------------------------------|
|   | Environmental                  | Social           | Economic                       |
| Y02C<br>(Capture, storage, sequestration or disposal of greenhouse Gas) | <b>0.1311</b><br><b>(43.2)</b> | 0.0833<br>(27.4) | 0.0893<br>(29.4)               |
| Y04S-10/58<br>(Economic aspect related to network operation)            | 0.0603<br>(19.3)               | 0.1154<br>(36.9) | <b>0.1373</b><br><b>(43.9)</b> |

온실가스의 처리와 관련된 특허 데이터는 환경 축 지속가능성이 가장 크게 계산되었으며, 네트워크 기술의 경제적 측면 관련 특허는 경제 축 지속가능성이 가장 크게 계산됨을 확인할 수 있다.

#### 5. 3D 프린팅 특허정보 대상 평가 결과

본 연구의 지속가능성 평가 방법이 특정 기술분야 또는 제품의 지속가능성 평가에 이용되는 예를 보이기 위해 FDM(FDM: fused deposition modeling, 이하 FDM) 방식의 3D 프린팅 기술분야에 출원된 특허 데이터를 대상으로 테스트하였다. 특허 데이터는 특허청에서 운영하는 특허정보 검색 사이트[24]를 이용하였으며, Fig. 2와 같이 키워드를 이용해 검색되 FDM 방식의 3D 프린팅 관련 특허를 활발하게 출원하고 있는 미국 Stratasys사의 특허정보가 검색되도록 보완하였다.

```

((Additive near2 (Manufactur* or Mfg* or Fab*)) or ((3D or 3-D or (three-dimension*) or (3 adj dimension*) or 3-dimension*) near2 print*) or (rapid* near2 Proto*)) and ((Fused near2 Deposition near2 Model*) or (Fused near2 Filament near2 Fab*) or (FDM) or (FFF))
+
(STRATASYS).AP.
```

Fig. 2. Search expression for patent data filed on FDM 3D Printing

미국 특허청에서 발행한 특허문서 246건, 유럽특허청에서 발행한 특허문서 73건 및 세계지적재산권기구에서 발행한 특허문서 129건을 추출하였으며, 검색된 특허문서에서 빈출되는 주요 IPC 분포는 Table 7과 같다.

다음으로, 본 연구의 지속가능성 평가 방법을 이용해 IPC별 지속가능성을 계산하였다. IPC별로 특허문서를 구분해 코퍼스(corpus)로 구성한 후 텍스트 마이닝을 통해 테스트 단어 집합을 만들었다. Table 7에 나열된 IPC 중 예시적으로 필라멘트와 관련된 IPC에 대한 지속가능성 계산 결과를 Fig. 3에 도시하였다. 예를 들어, 모집단을 형성하는 특허문서 중 IPC가 D01F로 부여된 특허문서는 6건인데, 이를 텍스트 마이닝을 통해 테스트 단어 집합으로 만든 후 본 연구의 평가 방법으로 지속가능성을 계산하면, TBL 환경, 사회, 경제 각 축별 비율은 각각 0.404, 0.367 및 0.228로 계산되는데, 그 비율을 Fig. 3에서 확인할 수 있다.

Table 7. IPCs of retrieved patent data filed on FDM 3D printing

| Tech. Area                 | IPC  | Description of IPC  | Quantity |
|----------------------------|------|---|----------|
| General Mechanism          | B28B | Shaping clay or other ceramic compositions, slag or mixtures containing cementitious material     | 12       |
|                            | B29C | Shaping or joining of plastics; shaping of material in a plastic state                            | 225      |
|                            | B33Y | Additive manufacturing  | 14       |
| Filament                   | B22F | Working metallic powder; manufacture of articles from metallic powder                             | 25       |
|                            | B65H | Handling thin or filamentary material   | 8        |
|                            | C08F | Macromolecular compounds obtained by re-actions only involving carbon-to-carbon unsaturated bonds | 6        |
|                            | C08J | General processes of compounding using organic macromolecular compounds                           | 11       |
|                            | C08L | Compositions of macromolecular compounds  | 13       |
|                            | D01F | Chemical features in the manufacture of man-made filaments, threads, fibers, bristles or ribbons  | 6        |
| Application of 3D printing | A61C | Dentistry; apparatus or methods for oral or dental hygiene  | 11       |
|                            | A61K | Preparations for medical, dental, or toilet purposes  | 11       |
|                            | G02B | Optical elements, systems, or apparatus   | 10       |
| Control of 3D printing     | G06F | Electric digital data processing  | 12       |

3D 프린팅에 사용되는 필라멘트에 대해 초기에 출원된 대부분의 특허문서는 B65H나 D01F에 속하는데, B65H나 D01F의 환경 축에 대한 지속가능성이 0.5 이라는 것을 알 수 있다. 최근 들어 출원되는 필라멘트 재

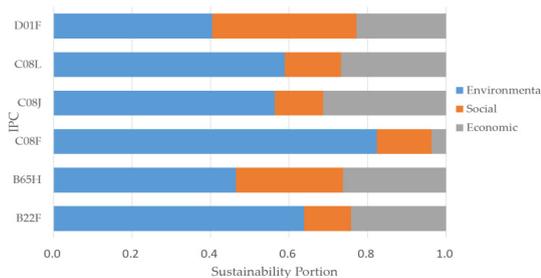


Fig. 3. Sustainability portion of IPCs in each axis for 3D printing filament

질과 관련된 특허문서는 B22F, C08J, C08L 등의 IPC가 부여되는데 세 개 IPC의 환경 축에 대한 지속가능성이 0.5 이상으로서 최대 0.63까지 커진 것을 확인할 수 있다. 이로부터 3D 프린팅의 필라멘트에 대한 기술은 환경 관련 지속가능성이 확대되는 방향으로 발전하고 있음을 유추할 수 있다.

다음으로, 본 연구에서 제시하는 지속가능성 평가 방법이 특정 기업의 지속가능성을 평가하는 데도 사용될 수 있다는 점을 보이기 위해 3D 프린팅의 대표적인 기업인 미국 Stratasys사가 출원한 특허 데이터를 대상으로 지속가능성을 계산하였다. 1989년부터 2017년까지 매년 Stratasys사가 출원한 FDM 방식의 3D 프린팅 특허들에 대해 기 계산된 IPC별 지속가능성 계산 결과를 기초로 각 축별 평균 지속가능성을 산정하였다. 예를 들어, 2005년도에 각각 B28B, B65H, D01F의 IPC가 부여된 3건의 특허가 출원되었다면, 위 3개 IPC의 각 축별 지속가능성 비율을 기초로 당해 연도의 지속가능성 비율이 결정된다.

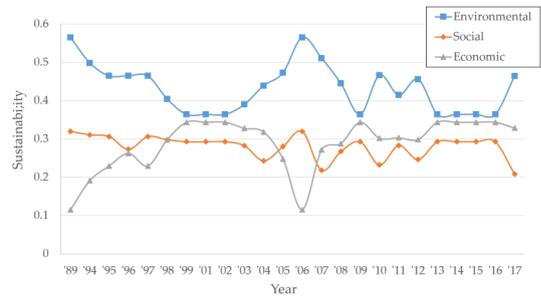


Fig. 4. Time series sustainability distribution of Stratasys' patent data

본 연구에서 제시하는 지속가능성 평가 방법에 따라 계산한 Stratasys사의 3D 프린팅 관련 지속가능성 변화는 Fig. 4와 같다. Stratasys는 전반적으로 환경 축 지속가능성이 높은 특허를 출원하고 있으며, 사회 축 지속가능성과 경제 축 지속가능성은 1997년 및 2006년에 역전되기도 했지만 전체적으로 경제 축 지속가능성이 높은 기술들을 출원하고 있음을 알 수 있다. 본 연구에서 제시하는 평가 방법으로 전 기간에 걸쳐 Stratasys사가 출원한 특허에 대해 지속가능성을 계산한 결과, Stratasys사는 평균적으로 환경적 지속가능성 43%, 사회적 지속가능성 26%, 경제적 지속가능성 32% 정도의 밸런스(balance)를 갖는 기업임을 확인할 수 있다.

## 6. 결론

본 연구는 텍스트 마이닝 기법을 이용해 TBL의 각 축 별 지속가능성을 자동으로 평가할 수 있는 새로운 정량적 지속가능성 평가 방법을 제안한다. 본 연구의 평가 방법은 텍스트 마이닝으로 사전을 구축하고 지속가능성 평가 대상 단어 집합과의 코사인 유사도를 이용해 지속가능성을 계산한다. 이에 더해, 본 연구는 사전에 포함된 단어의 관계성을 추가적으로 고려한다. 구글 검색 트렌드와 사회적 현상과는 상관관계가 있다는 연구에서 착안해 사전에 포함된 각 단어에 대해 구글 검색 데이터를 기초로 지속가능성 각 축 별 관계성을 부여한다.

지속가능성 평가 방법을 검증하기 위해 경향성이 알려진 문서 집합을 대상으로 테스트하였다. 환경적, 사회적, 경제적 지속가능성과 관련하여 각각 수집한 논문들에 적용한 결과, 환경 관련 논문은 환경 축 지속가능성이 가장 크게 계산되었으며, 사회 관련 논문은 사회 축 지속가능성이, 경제 관련 논문은 경제 축 지속가능성이 가장 크게 계산되었다. 환경 및 경제와 관련한 키워드가 포함되지 않도록 추출한 사회적 건강 관련 뉴스의 경우, 사회 축 지속가능성이 가장 크게 계산되었다. 본 연구가 제시하는 지속가능성 평가 방법이 특정 주제와 관련하여 모아진 문서 세트에 대해 비교적 타당하게 지속가능성을 평가하며, 각 축 별 지속가능성을 자동으로 정량적으로 계산할 수 있음을 알 수 있다. 또한, 본 연구의 평가 방법이 특허 데이터에 대해서도 타당한 결과를 낸다는 것을 확인하였으며, 특정 기술에 대한 지속가능성의 평가 가능성을 확인하였다.

본 연구의 평가 방법으로 특정 제품이나 기업에 대한 지속가능성 평가도 가능하다는 것을 보이기 위해 3D 프린팅과 관련하여 출원된 특허 데이터를 기초로 지속가능성을 계산하였다. 먼저, FDM 방식의 3D 프린팅 분야에 출원된 특허 데이터를 IPC를 기준으로 분류해 각 IPC에 해당하는 기술별 지속가능성을 계산하였으며, 이를 기초로 필라멘트 기술분야에 대한 지속가능성 현황 및 3D 프린팅 분야 대표 기업의 연도별 지속가능성 변화를 관찰하였다.

본 연구가 제안하는 지속가능성 평가 방법은 비정형 문서 분석 방법인 텍스트 마이닝 기법과 구글 검색 데이터를 기반으로 기업이 출원한 특허정보를 분석해 지속가능성 수준과 TBL 지속가능성 분포를 확인할 수 있는 일련의 특허정보 기반 평가 방법론으로, 종래의 수동 콘텐츠 분석 기법에 비해 인적 오류 가능성이 낮으며 다량의

분석을 단시간 내에 수시로 수행할 수 있다는 데에 의의가 있다. 이에 따라, 본 연구의 평가 방법을 기업이나 정부기관 등이 TBL의 각 축 별 지속가능성을 정량적으로 평가하고 주기적으로 모니터링 하면서 지속가능성을 높이는 활동을 지속하도록 촉진하는 일종의 도구로 활용할 수 있다.

## References

- [1] T. Kuhlman, J. Farrington, "What is Sustainability?", *Sustainability*, Vol.2, No.11, pp.3436-3448, 2010. DOI: <https://doi.org/10.3390/su2113436>
- [2] World Commission on Environment and Development (WCED), *Our Common Future*, Oxford University Press: New York, NY, USA, 1987.
- [3] J. Elkington, "Partnerships from Cannibals with Forks: The Triple Bottom Line of 21st-century Business", *Environmental Quality Management*, Vol.8, No.1, pp.37-51, 1998. DOI: <https://doi.org/10.1002/tqem.3310080106>
- [4] Global Reporting Initiative(GRI), *Global Reporting Initiative Standards on Economic, Environmental and Social Performance*, Available From: <http://www.globalreporting.org/standards/> (accessed March 23, 2022)
- [5] Corporate Register, Available From: <https://www.corporateregister.com/> (accessed March 23, 2022)
- [6] J. R. Modapothala, B. Issac, "Study of Economic, Environmental and Social Factors in Sustainability Reports using Text Mining and Bayesian Analysis", *Proceedings of 2009 IEEE Symposium on Industrial Electronics and Applications*, 2009. DOI: <https://doi.org/10.1109/ISIEA.2009.5356467>
- [7] R. Morris, "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations", *Journal of Management*, Vol.20, No.4, pp.903-931, 1994. DOI: <https://doi.org/10.1177/014920639402000410>
- [8] Q. Deng, M. Hine, S. Ji, S. Sur, "Building an Environmental Sustainability Dictionary for the IT Industry", *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp.950-959, 2017. DOI: <https://doi.org/10.24251/HICSS.2017.112>
- [9] L. Aaldering, R. Vliegthart, "Political Leaders and the Media. Can We Measure Political Leadership Images in Newspapers Using Computer-assisted Content Analysis?", *Quality & Quantity*, Vol.50, No.5, pp.1871-1905, 2015. DOI: <https://doi.org/10.1007/s11135-015-0242-9>
- [10] J. Kim, J. Choi, S. Park, D. Jang, "Patent Keyword

- Extraction for Sustainable Technology Management”, *Sustainability*, Vol.10, No.4, pp.1287-1304, 2018.  
DOI: <https://doi.org/10.3390/su10041287>
- [11] M. Asif, C. Searcy, P. d. Santos, D. Kensah, “A Review of Dutch Corporate Sustainable Development Reports”, *Corporate Social Responsibility and Environmental Management*, Vol.20, No.6, pp.321-339, 2013.  
DOI: <https://doi.org/10.1002/csr.1284>
- [12] I. Delai, S. Takahashi, “Corporate Sustainability in Emerging Markets: Insights from the Practices Reported by the Brazilian Retailers”, *Journal of Cleaner Production*, Vol.47, pp.211-221, 2013.  
DOI: <https://doi.org/10.1016/j.jclepro.2012.12.029>
- [13] W. J. Potter, D. Levine-Donnerstein, “Rethinking Validity and Reliability in Content Analysis”, *Journal of Applied Communication Research*, Vol.27, No.3, pp.258-284, 1999.  
DOI: <https://doi.org/10.1080/00909889909365539>
- [14] M. W. Bickel, “A New Approach to Semantic Sustainability Assessment: Text Mining Analysis Revealing via Transition Network Patterns in German Municipal Climate Action Plans”, *Energy, Sustainability and Society*, Vol.7, No.22, pp.1-25, 2017.  
DOI: <https://doi.org/10.1186/s13705-017-0125-0>
- [15] A. M. Shahi, B. I., J. R. Modapothala, “Intelligent Corporate Sustainability Report Scoring Solution Using Machine Learning Approach to Text Categorization”, *Proceedings of 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology*, pp.227-232, 2012.  
DOI: <https://doi.org/10.1109/STUDENT.2012.6408409>
- [16] A. M. Shahi, B. I., J. R. Modapothala, “Analysis of Supervised Text Classification Algorithms on Corporate Sustainability Reports”, *Proceedings of 2011 International Conference on Computer Science and Network Technology*, pp.96-100, 2011.  
DOI: <https://doi.org/10.1109/ICCSNT.2011.6181917>
- [17] M. Grimaldi, L. Cricelli, M. D. Giovanni, F. Rogo, “The Patent Portfolio Value Analysis: A New Framework to Leverage Patent Information for Strategic Technology Planning”, *Technological Forecasting and Social Change*, Vol.94, pp.286-302, 2015.  
DOI: <https://doi.org/10.1016/j.techfore.2014.10.013>
- [18] H. Ernst, S. Legler, U. Lichtenthaler, “Determinants of Patent Value: Insights from a Simulation Analysis”, *Technological Forecasting and Social Change*, Vol.77, No.1, pp.1-19, 2010.  
DOI: <https://doi.org/10.1016/j.techfore.2009.06.009>
- [19] MDPI, open access publishing, Available From: <https://www.mdpi.com/> (accessed March 23, 2022).
- [20] R Core Team (2015) R: A Language and Environment for Statistical Computing, Available From: <http://www.R-project.org> (accessed March 23, 2022).
- [21] H. C. Wang, Y. C. Chi, P. L. Hsin, “Constructing Patent Maps Using Text Mining to Sustainably Detect Potential Technological Opportunities”, *Sustainability*, Vol.10, No.10, pp.3729-3746, 2018.  
DOI: <https://doi.org/10.3390/su10103729>
- [22] D. M. Blei, A. Y. Ng, M. I. Jordan, “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, Vol.3, pp.993-1022, 2003.
- [23] X. Wan, “A Novel Document Similarity Measure Based on Earth Mover’s Distance”, *Information Sciences*, Vol.177, No.18, pp.3718-3730, 2007.  
DOI: <https://doi.org/10.1016/j.ins.2007.02.045>
- [24] Korea Intellectual Property Rights Information Service, Available From: <http://www.kipris.or.kr> (accessed March 23, 2022).

지 선 구(Seonkoo CHEE)

[정회원]



- 1998년 2월 : KAIST 항공우주공학과 (공학박사)
- 1998년 3월 ~ 2003년 6월 : (주)삼성 전자 책임연구원
- 2003년 7월 ~ 2014년 2월 : 특허청 특허심사관(변리사)
- 2015년 2월 ~ 현재 : 국립금오공과대학교 IT융합학과 교수

<관심분야>

산업재산권, 기술사업화, 산업재산 정책, 기술창업