

경영전문석사학위 논문

자연어 처리 방법론을 적용한  
국내 주요 기업의 ESG 인적자원관리  
인식 수준 평가

2023 년 2 월

서울과학종합대학원대학교

송 윤 석

자연어 처리 방법론을 적용한  
국내 주요 기업의 ESG 인적자원관리  
인식 수준 평가

지도교수 오 태 연

이 논문을 경영학 석사 학위논문으로 제출함

2023년 2월

서울과학종합대학원대학교

송윤석

송윤석의 석사 학위논문을 인준함

2023년 1월

위원장 장 중 호 (인)

위원 최 진 희 (인)

위원 오 태 연 (인)

## 초 록

최근 지속가능한 발전을 위한 기업과 투자자의 사회적 책임이 중요해지면서 세계적으로 많은 기업 및 정부기관, 유관단체들이 ESG 평가 정보를 활용하고 있다. ESG는 기업의 비재무적 요소인 환경(Environment), 사회(Social), 지배구조(Governance)를 뜻하는 용어이며, 최근 학계에서는 기업의 ESG 평가 항목에 대한 달성 수준과 기업의 성과를 나타내는 영역 간 관계에 대한 연구가 활발히 진행되고 있다.

국내 주요 기업들은 이와 같은 사회적 변화에 대응하고자 ESG 경영을 선언하고 입체적인 경영전략을 수립하고 있으며, 다양한 민간기관과 정부 당국은 ESG 경영 평가 수준을 평가하기 위한 영역을 제시하고 있다. 하지만, 기업의 성과와 관련된 연구는 주로 정량적으로 측정 가능한 관련 영역과 기업의 성과를 설명하는 연구 중심으로 진행되었으며, 비 정량적 측정 요소와 성과에 대한 연구는 아직까지 미비한 상황이다. 특히, 인적자원관리 영역에서는 인적자원관리의 질적 측면을 반영한 관련 영역을 설정하고 이를 ESG에 적용한 사례는 아직까지 찾아볼 수 없다.

이에 본 연구에서는 2019년 국제표준화기구(ISO, International Standardization Organization)에서 발표한 국제 표준 HR 보고서 가이드라인인 ISO 30414를 기준으로 국내 주요 기업의 ESG 또는 SDGs 보고서 상의 인적자원관리와 관련한 용어와 ISO 30414의 핵심용어와의 상관관계를 분석한다. 이를 통해 국내 대기업들의 ESG 중 사회(Social) 영역에서의 인적자원관리 인식 수준에 대한 평

가 방법론과 국제기준에 맞는 평가방향을 제시한다.

분석 방법론으로는 텍스트 마이닝을 위해 자연어처리 방법론을 적용한다. 2022년 11월 기준 국내 시가 총액 50대 기업의 ESG와 SDGs 영문 보고서 50건을 NLP(Natural Language Processing) 알고리즘인 문서 유사도 분석 기법을 활용하여 분석하였다. 분석 결과, 국내 대기업들이 발간한 ESG 리포트 상에서 ISO 30414 인적자원영역 내의 리더십, 생산성, 인력비용, 승계계획 항목에 대해서는 언급 횟수가 급감하거나 언급 자체가 없는 것으로 판명되었다. 이는 ESG 데이터에 관한 구체적인 공개 요구가 심화되고 있는 현실과 배치되는 상황을 드러내는 결과로서, 국내 주요 기업들의 대외 경쟁력 확보를 위한 관련 항목에 대한 관심이 필요한 상황을 반영한다.

본 연구의 학문적 의의는 ESG의 3대 요소 중 환경(Environment)과 지배구조(Governance)에 관한 연구에 비해 상대적으로 덜 주목받는 사회(Social) 요소의 인적자원관리 영역으로 연구의 범위를 확대시킨 부분에 있으며, 특히 자연어처리 분석 기법을 적용하여 국제 표준 가이드라인인 ISO 30414를 ESG 평가 기준으로 적용한 기초 연구로서 향후 관련 영역의 지표 개발에 활용 가능할 것으로 사료된다.

# **ABSTRACT**

## **Natural Language Processing on ESG Reports: The Case of Top Fifty Korean Enterprises Focused on The Area of Human Resource Management**

**Song, Yoon Seok**

**Seoul School of Integrated Sciences and Technologies**

Recently, as the social responsibility of enterprises and investors for sustainable development has become more important, many enterprises, government agencies, and related organizations around the world are using ESG evaluation information. ESG is the term that refers to company's non-financial factors, such as the environment, society, and governance. In the academic field, the research analyzing relationship between the company's ESG evaluation level and performance indicators is actively progressing.

Major domestic enterprises declaring ESG management and establishing various management strategies are following up this social change. Also, private organizations and government authorities are presenting indicators to evaluate the enterprises ESG management level. Most of studies handling the relation between ESG and corporate performance have been adopted quantitatively methodologies, and a only few studies based on non-quantitative approach method. Especially, in the field of human resource management, it is almost hard to find research cases applying qualitative indicator to evaluate ESG management level.

Therefore, this study focus on the similarity analysis between ESG(SDGs) reports of

Korea's top 50 companies and key terms related to human resources in ISO30414. ISO 030414 is the international standard human resources management report announced by ISO (International Standardization Organization). This report provides the guideline which grade the qualitative level of specific organizations' human resources management. Through text analyzing process, this study evaluates the recognition level of domestic enterprises' human resources management included social fields of ESG and suggest the insight for improvement. To induct statistical result, this study adopts document similarity analyzing process which is a kind of natural language processing (NLP).

As a result, in the ESG reports published by domestic Top 50 enterprises (based on KOSPI 200 market capitalization), specific items related human resources area within the ISO 30414 (leadership, productivity, labor cost and succession planning) are nothing or rarely mentioned. This result shows specific areas related human resources management which domestic top enterprises should pay attention to in terms of the international competitiveness. The academic significance of this study is that suggest new perspective for evaluating human resources management level of the enterprise in terms of ESG. In particular, it is meaningful that apply ISO 30414, an international standard guideline, as a significant ESG evaluation item, and adopt NLP as a new qualitative evaluation methodology.

Keywords: ESG(Environment, Social, Governance), HRM(Human Resources Management), ISO 30414, NLP(Natural Language Processing).

## 목 차

제 I 장 서론	1
제1절 연구 배경	1
제2절 연구 목적 및 구성	2
제 II 장 이론적 배경 및 선행연구 고찰	4
제1절 ESG 경영과 인적자원관리	4
제2절 ISO 30414 측정기준	6
제3절 국내외 ESG 관련 인적자원영역의 비교	6
제4절 선행연구 고찰	8
제 III 장 연구 방법	11
제1절 NLP (Natural Language Processing)	11
제2절 CountVectorization	13
제 IV 장 분석 및 연구 결과	16
제1절 데이터 분석 절차	16
제2절 데이터 분석 결과	20
제 V 장 결론 및 시사점	26
참고문헌	28

## 표 목 차

<표 1> 주요 가이드라인별 측정 영역 비교 . . . . .	7
<표 2> ISO 30414 인적자원영역별 언급 횟수 . . . . .	21
<표 3> ISO 30414 인적자원영역별 기초 통계량 . . . . .	22
<표 4> 개별기업별 인적자원영역 언급 횟수 . . . . .	24



# 제 I 장 서 론

## 제 1 절 연구 배경

ESG는 기업의 비재무적 요소인 환경(Environment), 사회(Social), 지배구조(Governance)를 축약한 용어이다. 이 용어는 2004년 유엔글로벌콤팩트(UNGC)와 20여 개 금융기관이 함께 발표한 “Who Cares Wins” 보고서에 처음 명시되었으며, 최근 사회책임투자(Socially responsible investing, 이하 SRI)<sup>1</sup>와 지속가능투자(Sustainable Development Goals, 이하 SDGs)의 관점에서 기업을 평가하기 위해 활용되고 있다. 기업의 ESG 성과를 활용한 투자 방식은 재무적 성과에 기초하여 투자하던 한계에서 벗어나 투자자들의 장기적 수익 창출에 도움을 주고, 기업 행동이 사회에 이익이 되도록 영향을 줄 수 있다는 점에서 그 중요성을 인정받고 있다.

관련 상황을 살펴보면 영국(2000년)을 시작으로 스웨덴, 독일, 캐나다, 벨기에, 프랑스 등이 자국 연기금을 중심으로 ESG 정보 공시 의무 체도를 도입해왔으며, 국내에서는 2021년 배포된 금융위원회 제시안에 따라 2025년부터 자산 2조 원 이상 코스피 상장사는 ESG 관련 지속가능한 보고서를 의무적으로 공시해야 한다. 국내의 경우 2020년 기준 코스피 상장사는 820여개 기업이며, 이 중 자산이 2조원을 상회하는 기업은 169개이다. 이들 중 ESG위원회를 이사회에 별도로 구성한 기업은 93개 기업(약 55%)에 불과하다.<sup>2</sup> 한편,

---

<sup>1</sup> 사회책임투자란 사회적·윤리적 가치를 반영하는 기업에 투자하는 방식으로서 투자 주체는 기업의 재무적 성과만을 판단하던 전통적 방식과 달리, 장기적 관점에서 기업 가치와 지속가능성에 영향을 주는 비재무적 요소를 의사결정에 반영한다.

<sup>2</sup> 연합뉴스 (2021.10.20.)

이 규정은 2030년부터 자산 규모와 상관없이 모든 코스피 상장사에 적용될 예정이다. 이처럼 ESG정보의 공개는 국내 기업에게 현실적으로 글로벌 경쟁력 확보와 지속가능성을 결정하는 중요한 이슈로 다가오고 있으며, 특히 ESG의 S(Social) 항목에는 인적자원관리와 관련된 사람/인간의 존엄과 권리, 평등, 안전, 다양성 등의 내용이 포함되어 있어, 이와 관련된 정보 공개 이슈 또한 중요하게 인식되고 있다.

## 제 2 절 연구 목적 및 구성

본 연구의 목적은 국내 50대 기업의 ESG 영역 중 사회(Social) 부분의 인적자원관리 주요 영역들이, 2019년 공개된 ISO 30414의 표준을 기준에 어느 정도 부합하는지 텍스트 분석 방법론을 적용하여 평가하고, 이를 통해 국내 주요기업들이 ESG 경영 내의 인적자원관리 영역에서 글로벌 스탠다드의 반영 정도를 파악하는 것이다.<sup>3</sup> ISO 30414의 영역이 ESG관련 보고서에서 많이 언급된 기업은 그 영역을 중요하게 인식하는 것으로 평가할 수 있다.

국제표준화기구(International Standardization Organization, 이하 ISO)에서 최초로 발표한 국제 인적자원관리의 표준 가이드인 ISO 30414는 인적자원관리에 관한 다양한 평가요소를 표준화된 양식으로 데이터화해서 제시함으로써 의사결정에 필요한 정보를 구조화 할 수 있다. 따라서, 향후 많은 기업들이 기업의 인적자원관리에 해당 인증을 활용하여 다양한 니즈에 대응하고 급격한 사회적 변화에 대응하기 위한 기준으로 정착될 가능성이 높으므로, 이에 대한 국내 주요 기업의 활용 현황을 분석하고 그 특징을 파악함으로써 우리 기업

---

<sup>3</sup> 2022년말 KOSPI200 시가총액기준

및 국가 경제의 글로벌 경쟁력을 강화하는 방안을 모색하는 기초 연구로서의 의의를 내포하고 있다.

본 논문은 총 5장으로 구성되어 있다. 1장에서 논문의 목적에 대하여 설명하고, 2장에서는 본 연구에서 ESG 경영 중 인적자원관리 수준의 분석 기준으로 채택한 국제 표준인 ISO 30414에 대해 알아본다. 3장에서는 자연어처리(Natural Language Processing, 이하 NLP) 방법론 중 단어 빈도수 기반의 임베딩(Embedding) 기법인 CountVectorizer에 대해 설명한다. 4장에서는 이를 적용하여 국내 50대 대기업 ESG 보고서의 데이터를 ISO 30414의 주요 텍스트와 비교 분석하고, 그 결과를 소개한다. 마지막 5장에서는 결과 및 향후 연구과제에 관해 기술한다.

## 제 II 장 이론적 배경 및 선행연구 고찰

### 제 1 절 ESG 경영과 인적자원관리<sup>4</sup>

ESG는 기업의 재무적 요소 이상으로 비재무적 요소인 환경(Environment), 사회(Social), 지배구조(Governance)를 중시하는 경영이념으로, 기업의 경제적 가치와 사회적 가치를 동시에 추구함으로써 지속 가능성을 담보할 수 있다는 철학을 내포하고 있다. UN에서 제시하는 ESG의 각 영역이 의미하는 바는, (1) 환경(Environment)의 경우 기후변화, 온실가스배출, 대기오염, 자원의 낭비 등을 의미하며, 이는 기업이 지구 환경에 미치는 영향을 말한다. (2)사회(Social)의 경우 기업이 사람을 어떻게 대하고 직원의 가치를 어떻게 평가하는지에 대한 내용으로서 건강과 안전, 인권, 다양성과 포용, 근로기준, 작업환경 등을 의미한다. (3)마지막으로 지배구조(Governance)의 경우 이사회가 다양하게 구성되어 있는지, 지배구조 및 감사 절차가 투명한지 등을 의미한다.

HRM (Human resources Management, 인적자원관리)은 ESG 경영의 사회(Social) 영역에서 다룬다. 2020년 8월 미국증권거래위원회(SEC, Securities and Exchange Commission)는 상장 회사에 대한 인적자본 공개를 의무화하였다. 과거와 달리 기업의 역량 및 자산은 재무적 요인만으로 설명이 어렵고, 비재무적 요인이 많은 영향을 미치고 있기 때문이다. 그리고 인적자본에 대한 정보는 이러한 비재무적 요인을 반영하는 대표적인 요소이다.

이재진(2021)에서는 미국 증권거래위원회(SEC)가 발표한 인적 자본에

---

<sup>4</sup> 본 절은 이재진(2021), HR 리더의 전략적 선택, ESG, 월간 인재경영 2021년 4월호, 제194호의 주요 내용을 인용

관한 보고 및 공시 규정과 국제표준화기구(ISO)에서 제공하는 인적자본 보고 가이드인 ISO 30414 는 향후 기업운영 방식에 크게 영향을 미칠 것으로 전망하였다. 그 이유는 이 두 기관이 제시하는 가이드가 비즈니스 시장 전체적으로 뜨겁게 논의되는 이슈인 ESG 트렌드와 일맥상통하기 때문이다. 특히 UN에서 제시하는 사회적(Social) 측면을 살펴보면, 다양성, 성 평등, 직원권리, 자선활동, 협력업체 직원에 대한 책임 등 HR과 직/간접적으로 연계된 항목들이 많이 있으며 이는 기업의 비 재무적인 요소의 중요성을 강조하는 ESG의 개념과 여러 측면에서 유사하게 적용할 수 있다고 판단하였다.

이재진 (2021)에서는 또한 미국 증권거래위원회(SEC)가 구체적인 명시를 의무화한 3가지 영역에 관하여 ISO 30414 표준이 제공하는 영역을 인력유치(Attraction), 인력개발(Development), 인력유지(Retention)로 구분하여 설명하였다. 인력유치는 인력 충원에 소요되는 시간, 중요한 포지션에 필요한 인력 충원 소요 시간, 내부 충원 비율, 핵심 포지션에 대한 내부 충원 비율을, 인력개발과 관련해서는 개발 및 훈련비용, 준법 및 윤리사항 교육을 이수한 지원의 비율을, 마지막으로 인력유지와 관련해서는 이직률을 중요 영역으로 언급하였다. ESG 정보 공개에는 인적자원과 관련된 정보가 포함되어 있고, 이는 곧 ISO 30414 내용과 상당부분 연계성을 가지고 있으므로, 국내 기업이 ISO 30414 가이드를 준용하여 인적관리자원 부분의 질을 향상시킨다면, 향후 ESG 정보 공개 요구 시 효과적인 대비가 가능할 것으로 판단하였다.

## 제 2 절 ISO 30414 측정기준<sup>5</sup>

조직의 내/외부 이해관계자를 위한 인적자원보고 표준 가이드라인인 ISO 30414는 국제 표준 모델은 2019년에 발표되었다. 이 표준은 최초로 국제적으로 채택된 HR 보고 가이드이며, 조직이 인적자원관리 관행 및 성과에 대해 보고할 수 있는 지침을 제공한다. ISO 30414는 11가지 핵심 영역과 이와 연계된 49가지의 지표로 구성되어 있다. 여기서 11가지 핵심 영역은 1) 규범 준수 및 윤리(Compliance and ethics), 2) 인력비용(Costs), 3) 다양성(Diversity), 4) 리더십(Leadership), 5) 조직문화(Organizational culture), 6) 조직의 건강, 안전 및 웰빙(Organizational health, safety and wellbeing), 7) 생산성(Productivity), 8) 채용, 유동성 및 이직률(Recruitment, mobility, and turnover), 9) 기술과 역량(Skills and capabilities), 10) 승계 계획(Succession planning), 11) 인력 가용성(Workforce availability)이다. 각 영역은 우선순위가 없이 동일한 수준으로 반영되며, 측정 기준의 특성에 따라 2~7개 지표로 구분하여 측정된다. ISO 30414는 단순히 HR 보고서 표준화 프레임워크가 아니라, HR 정보에 대한 공개 의무가 가속화되어 가는 현 시점에서 가장 활용 가능성이 높은 기준이 될 수 있다.

## 제 3 절 국내외 ESG 관련 인적자원영역의 비교

국내에서는 산업통상자원부를 중심으로 각 부처가 합동으로 K-ESG 영역을 개발하였고, 한국기업지배구조원(KSCGS)이 ESG평가 영역을 개발하였다. 하지

---

<sup>5</sup> 본 절은 이재진(2020), 최초의 국제 표준 HR보고서 가이드 - ISO 30414 월간 인재경영 2020년 10월호, 제188호의 주요 내용을 인용

만 인적자원영역에서는 최소한도의 영역만 제시하고 있다. SEC도 인적자원에 대한 중요성을 강조하고, 인적자원정보 공개를 의무화했으나, 정확한 측정 또는 목표를 지정하지 않았다. 대신 기업이 비즈니스의 특성에 따라 자신에게 가장 적합한 것을 결정할 수 있도록 하고 기본적인 영역만 공지하였다. K-ESG의 주요 측정기준은 총 7가지로 구분되며 이는 1) 노동-복리후생비, 2) 다양성 및 양성 평등, 3) 산업안전, 4) 생산성, 5) 노동-채용, 유지 및 이직률, 6) 노동-교육 훈련비, 7) 노동-정규직 비율이다. ISO 30414, SEC의 측정 영역별 내용은 <표 1>에 비교해 놓았다.

<표 1> 주요 가이드라인별 측정 영역 비교

주요 가이드라인	측정 영역
<b>K-ESG</b>	1) 노동-복리후생비 2) 다양성 및 양성 평등 3) 산업안전 4) 생산성 5) 노동-채용, 유지 및 이직률 6) 노동-교육 훈련비 7) 노동-정규직 비율
<b>ISO 30414</b>	1) 규범 준수 및 윤리 2) 인력비용 3) 다양성 4) 리더십 5) 조직문화 6) 조직의 건강, 안전 및 웰빙

	<ul style="list-style-type: none"> <li>7) 생산성</li> <li>8) 채용, 유동성 및 이직률</li> <li>9) 기술과 역량</li> <li>10) 승계계획</li> <li>11) 인력 가용성</li> </ul>
<b>SEC (JUST Capital Research)</b>	<ul style="list-style-type: none"> <li>1) 고용 및 노동 유형</li> <li>2) 임금, 보상 및 복리 후생</li> <li>3) 인력 다양성, 형평성 및 포용성</li> <li>4) 산업 보건 및 안전</li> <li>5) 직업 안정성</li> <li>6) 훈련 및 교육</li> </ul>

세 가지 영역을 비교해보면 ISO 30414의 기준이 가장 구체적이고 세부적인 측정 기준 항목을 명시하고 있다. 결과적으로 ISO 30414 적용 시 기업의 인적자원관리 수준을 가장 객관적으로 평가할 수 있으며, 향후 ESG 인적자원 관리 평가항목 설정 과정에서도 이러한 요소가 반드시 반영되어야 할 것으로 판단된다.

#### 제 4 절 선행연구 고찰

기존의 국내 ESG 연구들은 이론적 연구에 치중하고 있거나, 환경(Environment)과 지배구조(Governance)에 초점이 맞추어져 있고, 사회(Social) 부문에서는 인적자원관리와 관련한 사항이 중요하나, 평가는 주로 대외적 활동인 CSR(Corporate Social Responsibility) 분야에서 진행되고 있다. 관련 주요 연



구의 내용은 아래와 같다.

조대형 (2021)에서는 전세계적으로 증가하고 있는 ESG에 대한 높은 관심을 반영하여 글로벌 ESG 추진현황과 해외 주요 기업의 ESG 경영에 대한 실제 사례를 분석하였다. 해당 연구는 기업의 지속가능한 성장을 위해 ESG가 필수불가결한 패러다임이라는 것을 실제 사례를 통해 주장하였고 특히, 해외 주요 기업의 사례를 통해 ESG가 기업의 단기적 수익 뿐만 아니라 장기적인 경쟁력 강화에도 긍정적인 영향을 미친다는 사실을 확인하였다.

천상은, 박희태 (2021)는 기업의 비재무적 특성을 강조하는 ESG 경영의 활성화에 있어 인적자원관리의 역할을 강조하면서 해당 부분에 대한 관련 연구 동향 및 국내 대표적 ESG 평가기관들의 영역을 비교 분석하여 인적자원관리 영역에서의 연구확장을 도모하였다. 아울러, ESG 경영 평가에서 인적자원관리 분야의 중요성을 재정립하고 해당 분야가 내포하고 있는 개별 기업 간 특성에 대한 차별성을 강조하면서 향후 ESG 경영에서 인적자원관리 분야의 중요성이 더욱 증대될 것임을 주장하였다.

안태욱 외 (2021)는 날로 급증하고 있는 ESG 정보공개에 대응하기 위한 기업과 투자자 간의 인식에 대한 접근방법으로서 연관규칙 분석을 통한 ESG 우려사안 키워드 도출에 관해 연구하였으며 이를 통해 투자자들이 파악하고 있는 반ESG 활동이 어떤 것인가를 분석하였다. 연구결과의 도출을 위해 2013년부터 2018년까지의 기간 동안 294개 상장기업과 관련되어 미디어에 노출된 반ESG 활동 기사를 대상으로 연관규칙 분석방법을 활용하여 ESG 우려사안 키워드를 추출하였다. 이를 통해 지배구조관련 키워드가 최상위에 위치하고 있음을 확인하였으며 추상적인 개념의 ESG 우려사안을 쉽게 파악할 수 있도록 하여 다양한 이해관계자들이 활용할 수 있도록 연구를 구성한 부분에서 의의를 갖는다.

임효진 (2021)은 국제적 기업 평가의 기준으로 부상하고 있는 ESG의 위상에 대응하기 위한 연구의 일환으로서 중소기업의 ESG 이슈 분석의 연구를 수행하였다. 이를 위해 중소기업과 ESG를 핵심 단어로 투입하여 뉴스기사 텍스트 마이닝 기법을 활용한 분석을 실시하였고, 환경, 사회, 지배구조 등 각각의 ESG 분야에 대한 핵심 화두를 추출하여 현재 중소기업이 직면하고 있는 핵심 이슈를 도출해 내었다. 특히, 텍스트 마이닝 기법을 활용하여 실질적이고 구체적인 중소기업 관련 이슈를 특정하고 이를 통해 정부 및 산업계에서 활용 가능한 정책적 대안을 이끌어 내기 위한 기초자료로서 활용될 수 있는 가능성을 제시하였다.

위에서 살펴본 바와 같이, ESG에 대한 전세계적인 관심에 걸맞게 다양한 주제의 연구가 진행되었고 그 범위가 점차 확대되어가고 있음에도 불구하고, 인적자원관리 측면에서 ESG를 평가하고 연관성을 도출해 나가는 연구는 상대적으로 부족한 실정이다. 아울러, 다양한 머신러닝 기법 또는 텍스트 마이닝 기법을 활용하여 핵심 단어나 주제어를 특정하고 이를 토대로 정책적 대안을 제시하는 연구방식이 점차 증가하고 있는 추세를 반영하여 본 연구에서는 ESG 경영에 있어서 인적자원관리 영역의 중요성과 이를 확인하기 위한 도구로서 NLP 기법을 적용한 분석을 진행함으로써 ESG 연구의 외연을 확장하고자 한다.

## 제 III 장 연구방법

### 제 1 절 NLP (Natural Language Processing)

자연어 처리(Natural Language Processing, NLP) 분야는 1990년대 초까지 규칙(Rule) 기반의 기계번역이 주를 이뤘다. 그러나 많은 양의 자연어 단어와 문장에 대해 일일이 규칙을 설정하고, 처리하는 것은 시간과 노력 대비 결과물이 좋지 않았다. 이후 컴퓨팅 파워가 점차 증가함에 따라 기계학습 알고리즘을 사용한 통계 기반의 자연어 처리가 도입되었다. 그리고 현대에는 자연어 처리에 관한 다양한 연구와 압도적으로 증가한 계산량에 힘입어 신경망 기반의 자연어 처리가 주를 이루게 되었다. 자연어 처리는 세부적으로 많은 기술 및 애플리케이션을 포괄한다. 자연어 처리 기술을 이용하면 기초적인 텍스트 전처리에서부터, 텍스트/음성 변환, 품사 태깅, 엔티티 인식, 감정 분석, 관계 추출과 같은 일차적 작업과 자동 요약, 기계번역, 챗봇, 자연어 이해, 질의응답 등 고도화된 애플리케이션까지 수행할 수 있다.

자연어 처리(NLP) 작업에는 다음을 포함하되 이에 국한되지 않는 여러 범주가 있다. ① 텍스트 분류(Text Classification): 주어진 텍스트에 미리 정의된 범주 또는 레이블을 지정한다. 예를 들어 스팸 감지, 감정 분석 및 주제 분류가 있다. ② 언어 번역(Language Translation): 한 언어에서 다른 언어로 텍스트를 번역한다. ③ 텍스트 요약(Text Summarization): 주요 정보를 유지하면서 주어진 텍스트의 짧은 버전을 생성한다. ④ 명명된 엔티티 인식(Named Entity Recognition): 주어진 텍스트에서 사람, 조직 및 위치와 같은 명명된 엔티티를 식별하고 분류한다. ⑤ 감정 분석(Sentiment Analysis): 주어진 텍스트에 표현된 의견이나 감성을 결정한다. ⑥ 텍스트 생성(Text Generation): 스타일이나 내용이 주어진 텍스트와 유사한 새 텍스트를 생성

한다. ⑦ 텍스트 음성 변환 및 음성 인식(Text-to-Speech and Speech Recognition): 서면 텍스트를 음성 단어로 변환하고 음성 단어를 서면 텍스트로 인식한다. ⑧ 언어 모델링(Language Modeling): 주어진 텍스트에서 단어의 확률 분포를 모델링한다. ⑨ 대화 시스템 및 챗봇(Dialogue System and Chatbots): 사용자와 사람처럼 대화할 수 있는 시스템 구축한다. ⑩ 텍스트 함의 및 의역 감지(Textual Entailment and Paraphrase Detection): 두 텍스트 사이의 관계를 식별하고 의역을 감지한다.

NLP 전처리 기법에는 ① Tokenization: 텍스트를 개별 단어나 구문으로 분할하는 것이다. ② Stop Word Removal: 불용어 제거, 일반적으로 의미가 없는 단어들을 제거하는 것이다. ③ Lemmatization and Stemming: 어근 추출과 원형 복원, 단어를 일반적인 형태로 바꾸는 것이다. ④ Part-of-Speech Tagging: 품사 태깅, 각 단어의 품사를 판별하는 것이다. ⑤ Named Entity Recognition: 이름 부여된 개체 인식, 사람, 조직, 장소 등을 인식하는 것이다. ⑥ Text Normalization: 텍스트 정규화, 영문 대소문자, 숫자, 특수 기호 등을 정규화 하는 것이다. ⑦ Text cleaning: 텍스트 정리, 각종 오류와 잘못된 표기를 정리하는 것이다. ⑧ Text to numerical representation: 텍스트를 수치적 표현으로 변환, 예를 들면 CountVectorization, TF-IDF(Term Frequency - Inverse Document Frequency) 등이 있다.

NLP의 일반적인 순서는 다음과 같다. ① Data Collection: 사용할 텍스트 데이터를 수집한다. ② Data Preprocessing: 수집한 데이터를 전처리 한다. 일반적으로 tokenization, stop word removal, lemmatization and stemming, Part-of-Speech tagging, named entities recognition, text normalization, text cleaning, Text to numerical representation 등을 수행한다. ③ Exploratory Data Analysis (EDA): 전처리된 데이터를 분석하여 이해한다. EDA를 통해 데이터의 분포, 어휘, 언어 특징 등을 이해할 수 있다. ④

Modeling: 모델을 선택하고 학습시킨다. NLP에서는 다양한 모델을 사용할 수 있다. 예를 들면 기계 학습 모델 (SVM, Random Forest, Neural Network 등), rule-based 모델, probabilistic model 등을 사용할 수 있다. ⑤ Evaluation: 학습시킨다.

전처리가 완료된 텍스트는 연산을 위해 벡터화되는 임베딩을 거친다. 자연어 처리(NLP)에 임베딩은 단어, 구 또는 문서를 고차원 공간에서 밀도가 높은 저차원 벡터로 나타내는 데 사용되는 기술이다. 임베딩 이면의 아이디어는 의미적으로 유사한 단어가 서로 가까운 연속 벡터 공간에서 단어의 의미와 컨텍스트를 캡처하는 것이다. 임베딩은 종종 텍스트 데이터에 대한 기계 학습 모델을 교육하기 전에 전처리 단계로 사용된다. 결과 임베딩 벡터는 텍스트 분류, 언어 번역, 감정 분석 및 명명된 엔터티 인식과 같은 다양한 NLP 작업에 대한 입력으로 사용할 수 있다. GloVe와 같은 사전 훈련된 임베딩을 사용하거나 word2vec 또는 GloVe와 같은 신경망을 사용하여 처음부터 훈련 임베딩을 사용하는 등 단어 임베딩을 얻는 방법에는 여러 가지가 있다. 또 다른 방법은 다양한 NLP 작업을 위해 미세 조정할 수 있는 대규모 텍스트 데이터 코퍼스에 대해 사전 훈련된 변환기 기반 모델인 BERT를 사용하는 것이다.

## 제 2 절 CountVectorization

데이터를 분석함에 있어서 특징을 추출하는 것은 무척 중요한 일이다. 특징은 데이터 분석의 입력 값으로 사용될 만큼의 가치를 지닌 비교적 간단한 형태의 데이터인데, 특징 추출을 잘하면 분석의 중요도가 낮은 데이터를 제거하고 변환과 노이즈에 강한 분석을 할 수 있다. 특히 영상 데이터, 자연어 데이터와 같은 분석이 어려운 데이터를 비교와 해석에 필요한 형태로 변형시

켜서 처리가 가능해지므로 특징 추출은 특정 영역의 데이터 분석을 위해 반드시 필요한 과정이다. 특징은 데이터 종류에 따라 다른 특징이 존재하므로 적절한 방식으로 추출해야 한다. 텍스트 특징 추출의 대표적인 방법은 아래와 같다.

- 1) **CountVectorizer**: 각 텍스트에서 단어 출현 횟수를 계수한 벡터로, 그 형태(shape)은  $[n \times f]$  로 나타낼 수 있는데(실제로는 벡터가 아닌 행렬이라고 표현하는 것이 정확함), 여기서  $n$ 은 문서(또는 문장)의 개수,  $f$ 은 feature, 즉 단어의 개수가 되며, 벡터 내의 위치의 값은 해당 단어의 출현 횟수가 된다. 예를 들면, 문서(또는 문장)의 개수가 5, 모든 문서(또는 문장)에서 등장하는 단어의 수가 10일 때 **CountVector**는  $[5 \times 10]$ 의 형태가 된다(5 rows by 10 columns), 단 이때 출현횟수가 너무 적은 단어나 너무 많이 등장하는 단어(예: she, they 등)은 벡터화하기 전에 제거해 주어야하는 하는 번거로움이 있다.
- 2) **TfidfVectorizer**: TF-IDF<sup>6</sup> 값을 사용하여 **CountVectorizer**의 단점을 보완<sup>7</sup>한 방법으로, 위의 **CountVector**와 비교하여 중요도에 비해 빈도수가 너무 많은 단어에 대한 가중치를 줄여주는 장점이 있다. TF-IDF를 구하는 방식은 아래와 같다.
  - $TF-IDF = TF(t, d) \times \log(1 / IDF(t))$ 
    - $TF(t, d)$  : 특정 문서  $d$ 에서 특정 단어  $t$ 가 등장하는 횟수
    - $IDF(t)$  : 특정단어  $t$ 가 등장한 문서의 수:  $t$ 라는 단어가 모든 문서에서 자주 등장하면 불필요한 단어라고 간주함
- 3) **HashingVectorizer**: **CountVectorizer**에서 해시 함수를 사용하여 속도 향상:

---

<sup>6</sup> Term Frequency-Inverse Document Frequency, 여러 문서에서 어떤 단어가 특정 문서에서 중요한지 보여주는 통계적 수치

<sup>7</sup> <https://wiserloner.tistory.com/917>의 내용을 참조하였음.

HashingVectorizer는 처리속도의 향상이라는 점을 제외하면 연관성 분석을 목적으로 하는 분야에서는 사용하기 어려움

CountVectorizer의 특징은 텍스트에서 단위 별 등장횟수를 계산(카운팅)하여 수치벡터화 하는 것이다. 단위는 문서 단위, 문장 단위, 단어 단위를 정할 수 있다. 텍스트 카운팅은 먼저 단어 사전 벡터를 만들고, 카운팅 할 문장을 확인하며 그 단어 사전의 횟수를 카운팅하는 것이다. 본 논문에서는 위의 세가지 방식 중 기본평가 항목 텍스트에 대한 기업별 보고서의 유사성이 아닌 일치성을 평가하는 것에 초점을 맞추었으므로, 유사도 평가 알고리즘이 아닌 일치성 평가에 유리한 CountVectorization을 변형하여 사용하였다.

## 제 IV 장 분석 및 연구 결과

### 제 1 절 데이터 분석 절차

본 논문에서는 분석을 위한 데이터셋을 마련하기 위해 2022년 11월 기준 국내 시가총액 상위 50대 기업의 ESG와 SDGs 영문 보고서 50건을 각 회사의 홈페이지에 접속하여 수집하였다. 이후 ISO 30414 Human Capital Reporting Process 텍스트 분석 절차는 다음과 같은 과정을 걸쳐 수행되었다. 가장 우선적으로 기본절차에 해당하는 과정을 순서대로 나열하면, 첫째, 기준평가항목 텍스트인 (ISO\_30414\_2018(en).pdf)에서 평가항목에 해당하는 텍스트를 추출하고 둘째, NLP 기법을 적용해 해당 텍스트와 타 텍스트와의 유사성을 비교하기 용이하도록 토큰화(tokenization) 작업을 수행하였다. 셋째, 각 기업별 ESG 지침 및 보고서를 수집하고 넷째, 수집된 각 보고서를 위와 동일한 방법으로 가공한 후 다섯째, 기본평가항목 텍스트를 기준으로 기업별 보고서가 얼마나 유사 또는 일치하는지를 평가하였다. 여섯째로 위의 텍스트 유사/일치서 평가에는 다음과 같이 검토가능한 기술적 방법이 있는데, Levenshtein, V. I. (1966)에서 언급한 문장(sentence)별 Levenshtein Distance 비교 방식 또는, 각 텍스트를 TF-IDF Vectorizer<sup>8</sup> 로 변형 후 코사인 유사도(Cosine Similarity)로 비교하는 방식, 각 텍스트를 CountVectorizer<sup>9</sup>로 변형 후 코사인

---

<sup>8</sup> 해당 방식에 대한 설명은 다음 웹사이트에 설명되어 있음 ([https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.TfidfVectorizer.html#sklearn.feature\\_extraction.text.TfidfVectorizer](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#sklearn.feature_extraction.text.TfidfVectorizer)).

<sup>9</sup> CountVectorizer는 [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.CountVectorizer.html#sklearn.feature\\_extraction.text.CountVectorizer](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html#sklearn.feature_extraction.text.CountVectorizer)의 방식을 참고함.



유사도로 비교하는 방식과 각 테스트를 BM25<sup>10</sup> 방식으로 비교하는 형태, 그리고, TextRank<sup>11</sup> 알고리즘으로 각 텍스트의 주요 문장 Top-N개를 추출하여 일치성을 비교하는 방식으로 구분할 수 있다. 마지막으로 기업별로 해당 항목별 일치성 수치에 대한 분석으로 기본절차가 수행되었다.

본 논문에서는 위의 기본평가 항목 텍스트에 대한 기업별 보고서의 유사성이 아닌 일치성을 평가하는 것에 초점을 맞추었으므로, 유사도 평가 알고리즘이 아닌 일치성 평가에 유리한 CountVectorization을 변형하여 사용하였다. 즉, 기본평가 항목의 평가 항목별 키워드를 추출하여 해당 키워드가 각 기업별 보고서에 얼마나 등장하는지를 평가하였다. 이를 위하여 아래와 같이 2차에 걸쳐 평가를 시도하였으며, 아래 두번째 방식을 채택하였다. 여기서 사용된 자연어처리(Natural Language Processing, 이하 NLP) 기법 및 도구는 다음과 같다. 먼저, Python fitz 패키지를 통해 PDF파일에서 텍스트를 추출하였고, Python nltk(natural language tool kit) 패키지 중 1) nltk.sent\_tokenize 로 텍스트를 문장 단위로 분리하였고, 2) nltk.word\_tokenize 는 문장을 토큰 단위로 분리하였으며, nltk.stem.WordNetLemmatizer로 어원을 복원하고, nltk.pos\_tag로 텍스트에 대한 형태소분석(morphological analysis)을 실시하였다. 정규표현식(regular expression)을 통해 텍스트 분석과정에서 불필요한 공백/문장부호(punctuation mark) 등을 제거하였고, Python text\_to\_num 패키지의 text\_to\_num.alpha2digit으로 텍스트 내 품사 중 수사(numeral)를 숫자로 변환하는 작업을 수행하였다. 차수 별 분석 수행 내용은 아래와 같다.

(1차) 평가항목 분석 방식을 정리하면 다음과 같다. 첫째로, 기준평가항목 텍스트 파일(ISO\_30414\_2018(en).pdf)에서 텍스트를 추출하고

---

<sup>10</sup> 해당 방식은 Trotman et al. (2014)을 참고하였음.

<sup>11</sup> Mihalcea, R., Tarau, P. (2004)에 수록된 내용을 참고하였음.

둘째, 각 텍스트에서 평가항목 부분에 대해 자연어 처리기법으로 필요항목을 추출하였다. 그 과정에서 텍스트 내 맞춤부호(punctuation mark)를 제거 후 문장(sentence) 단위로 배열하였으며, 각 문장에 대해 형태소분석을 실시, 텍스트를 토큰(token)과 품사의 쌍으로 분리하는 작업을 수행하였다. 각 문장 중 숫자부분 및 숫자로 이루어진 범위를 표준화하는 작업을 진행하였는데 예를 들어, “fifty to two hundred and three” 을 “50-203” 으로 표준화하는 것과 같은 방식으로 설명할 수 있다. 또한, 숫자부분 외의 토큰 중 명사(nouns)에 해당하는 토큰만 추출한 후, 표준화된 분석을 위해 추출된 명사에 대한 어원복원(word stemming)을 수행하였는데 이는, children을 child로, employees를 employee등으로 원형 복원하는 것을 의미한다. 그 후, 각 평가항목에 포함된 명사 및 숫자(또는 숫자 범위)를 집합(set)으로 구성하였다. 셋째, 각 기업의 ESG 보고서의 전체 텍스트에 대해 위의 “둘째” 과정을 반복하여 처리하였고 넷째, 반복적인 분석을 위해 위의 “둘째 ~ 셋째” 과정의 결과를 Python의 피클(.pickle) 파일로 저장하였으며, 다섯째, 위의 “둘째”과정에서 얻은 각 평가항목별 키워드 집합에 대해 “셋째”과정에서 추출한 각 기업의 보고서 분석 결과물에 대해 일치하는지 여부를 교집합(intersection) 연산을 통해 집계하였고 마지막 여섯째로 각 기업별 집계 현황을 비교하여 분석하였다.

(2차) 평가항목 분석 방식 역시 순서대로 나열하면 위의 “(1차) 평가항목 분석 방식”과 대부분의 과정이 유사하나 2차 과정에서는 좀 더 엄밀한 일치성을 확인하기 위해 항목별 키워드를 집합이 아닌 키워드 문자열로 비교하는 방식을 선택하였다. 위의 “(1차) 평가항목 분석 방식”에서 추출한 텍스트를 기준으로 필수적인 키워드를 선택하였는데 이때, 평가항목에 필요한 필수 키워드는 한 단어가 아닌 어구(phrase) 단위로, 필요할 경우 복수 개의 어구를 채택하였다. 또한, 각 텍스트에서 평가항목 부분에 대해 자연어

처리방법으로 필요항목을 추출하였는데, 이 과정은 위의 “(1차) 평가항목 분석 방식”의 과정과 대부분 동일하나 최종 결과물의 형태가 “집합(set)”의 배열이 아닌, 표준화된 어구의 배열로 되도록 한 점이 다르다. 이후, 각 기업의 ESG 보고서의 전체 텍스트에 대해 위의 “셋째”과정을 반복하여 처리하였고, 반복적인 분석을 위해 위의 “셋째”, “넷째”과정의 결과를 Python의 피클(.pickle)파일로 저장한 후, 위의 “셋째”과정에서 얻은 각 평가항목별 키워드 배열에 대해 “셋째”과정에서 추출한 각 기업의 보고서 분석 결과물에 대해 일치하는지 여부를 문자열 비교 연산을 통해 집계한 후, 각 기업별 집계 현황을 비교하여 분석하는 방식을 사용하였다.

(3차) 위의 “2차” 과정에서는 평가원본 텍스트와 각 기업에 대한 보고서를 비교하는 과정에서 “전치사(preposition)”까지 포함한 문구를 그대로 비교하였다. 이 경우, 해당 기업에서 적절한 평가항목을 제시하였음에도 불구하고 적절히 평가되지 않을 가능성이 존재한다. 예를 들면 “Quality per hire” 대신 “Quality of hire”를 사용한 경우라 할 수 있다. 따라서, “3차” 과정에서는 텍스트를 비교할 때 형태소 분석(morphological analysis)을 통해 형용사와 명사에 해당하는 품사를 제외한 단어를 제거한 후 비교하였다. 이 때, nltk를 사용하여 품사를 추출할 때 채택된 품사목록은 아래와 같다.

- JJ: (일반 형용사, Adjective, 예: ‘big’ )
- JJR: (비교급 형용사, Comparative Adjective, 예: ‘bigger’ )
- JJS: (최상급 형용사, Superlative Adjective, 예: ‘biggest’ )
- NN: (일반 명사, Noun, 예: ‘child’ )

- NNS: (복수 명사, Plural Noun<sup>12</sup>, 예: ‘children’)
- NNP: (고유 명사, Proper Noun, 예: ‘Harrison’)
- NNPS: (고유 복수 명사, Plural Proper Noun, 예: ‘Americans’)

## 제 2 절 데이터 분석 결과

총 50개의 대상 기업 중 최종적으로는 46개사의 자료가 활용되었는데, 3개회사는 ESG 리포트를 공개하지 않았고 1개사는 파일에서 text를 추출하지 못해 사용이 불가능하였다. 아래 <표 2>에서 보듯이 인적자원영역이 많이 언급된 순서대로 나열하면 조직문화, 규범준수 및 윤리, 다양성, 기술과 역량에 대해서는 국내 주요기업의 ESG 리포트에서 언급 횟수가 많으나, 기타 조직의 건강, 안전 및 웰빙, 인력가용성, 채용, 이동 및 이직률, 리더십, 생산성, 인력비용, 승계계획에 대한 영역에서는 매우 낮거나 언급이 없는 것을 알 수 있다. 특히, 채용, 이동 및 이직률, 리더십, 생산성, 인력비용, 승계계획에 영역들의 언급횟수는 100회 이하로 급격하게 감소하는 모습을 파악할 수 있으며 승계계획의 경우에는 언급 횟수가 아예 없는 것으로 나타남으로써 국내 대기업들의 승계계획에 관한 소극적인 정보공개 태도를 확인할 수 있었다.

---

<sup>12</sup> 이 경우 본 논문에서는 ‘children’ → ‘child’와 같이 stemming 처리를 하여 단어의 원형을 사용하였음

<표 2> ISO 30414 인적자원영역별 언급 횟수

순서	ISO 30414 인적자원영역	언급횟수
1	조직문화	2805
2	규범준수 및 윤리	2720
3	다양성	1356
4	기술과 역량	1011
5	조직의 건강, 안전 및 웰빙	448
6	인력 가용성	227
7	채용, 이동 및 이직률	79
8	리더십	13
9	생산성	9
10	인력비용	6
11	승계계획	0

ISO 30414의 11개 영역에 관한 전반적인 기초 통계량은 <표 3>에 정리되어 있다. 분석 대상 기업 수는 총 46개사이고, 11개 영역의 평균 언급 횟수는 조직문화가 48회로 가장 많으며 다음으로 규범준수 및 윤리(34회), 다양성(25회), 기술과 역량(12회), 조직의 건강, 안전 및 웰빙(8회), 인력 가용성(3회), 채용, 이동 및 이직률(3회) 순으로 나타났다. 하지만, 조직의 실질적 지속성을 알아볼 수 있는 리더십, 생산성, 인력비용, 승계계획 이 네가지 영역의 평균값은 0회로 도출되었다.

<표 3> ISO 30414 인적자원영역별 기초 통계량

	규범 준수 및 윤리	인 력 비 용	다 양 성	리 더 십	조 직 문 화	조직의 건강, 안전 및 웰 빙	생산 성	채용, 이동 및 이직 률	기술 과 역량	승 계 계 획	인 력 가 용 성
<b>count</b>	46	46	46	46	46	46	46	46	46	46	46
<b>mean</b>	59	0	29	0	61	10	0	2	22	0	5
<b>std</b>	34	0	25	0	48	8	0	3	12	0	3
<b>min</b>	0	0	0	0	0	0	0	0	0	0	0
<b>25%</b>	35	0	14	0	28	4	0	0	15	0	3
<b>50%</b>	52	0	21	0	43	8	0	2	21	0	5
<b>75%</b>	84	0	38	1	83	13	0	2	29	0	6
<b>max</b>	146	2	111	1	196	31	2	17	51	0	12

개별 대상기업들의 인적자원영역의 언급 횟수는 <표 4>에서 살펴볼 수 있다. 조사 대상 기업 중 인적자원영역을 가장 많이 언급한 상위 10위 중 1위 기업은 포스코 로서 411회의 언급 횟수를 기록하였다. 그 다음으로는 카카오 및 카카오페이(408회)<sup>13</sup>, KT&G(326회), 현대글로비스(309회), 삼성SDS(281회), SK(280회), 현대중공업(280회), SKT(277회), 하나금융그룹(264회) 이다.

포스코의 경우, 규범준수 및 윤리에 대한 언급이 가장 많았고 다양성과 조직문화, 기술과 역량의 순으로 언급 횟수가 집계되었다. 카카오 및 카카오페이의 경우는 조직문화가 가장 많이 언급되었고, 다음으로는 규범 준수 및 윤리, 다양성의 순을 보였으며, KT&G는 규범준수 및 윤리, 다양성, 조직문화의 순으로 분석되었다.

<sup>13</sup> 두 기업의 리포트 내용 확인 결과 동일한 내용으로 발간된 되었다.

개별항목별로는 조직문화가 총 2805회 언급되어 가장 많이 표출된 영역이었으며 규범준수 및 윤리(2720회), 다양성(1356회)가 상대적으로 많이 언급된 것으로 분석되어, 우리나라 대기업들의 인적자원영역에서의 관심도가 규범과 윤리를 중시하면서도 조직문화와 다양성을 합리적으로 관리하기 위한 노력을 하고 있음을 반영하는 것으로 사료된다. 반면, 채용, 이동 및 이직률(79회), 리더십(13회), 생산성(9회), 인력비용(6회), 승계계획(0회) 등은 언급횟수가 상대적으로 낮게 나타났는데, 이는 국내 대기업들의 인적자원관리 분야에서 해당 영역이 심도 있게 다뤄지지 못하고 있음을 반영하는 것으로 파악할 수 있다. 특히, 기술과 역량(21회), 인력비용(6회) 등의 영역은 인적자원관리보다는 기업의 기타관리영역에서 다뤄지고 있음을 유추할 수 있으며 모든 기업에서 전혀 언급되지 않는 승계계획에 관한 내용은 기업특성이 이와 관련이 없거나 이와 관련해 소극적으로 대응하는 기업의 태도를 반영하는 것으로 유추할 수 있다.

<표 4> 개별기업별 인적자원영역 언급 횟수

	총 계	규 범 준 수 및 윤 리	인 력 비 용	다 양 성	리 더 십	조 직 문 화	조 직 의 건 강, 안 전, 웰 빙	생 산 성	채 용, 이 동 및 퇴 사 율	기 술 과 역 량	승 계 계 획	인 력 가 용 성
<b>Total number</b>		2720	6	1356	13	2805	448	9	79	1011	0	227
LGChemical_SDG_2021	92	8	0	20	0	42	3	1	0	16	0	2
SamsungSDS_SDG_2022	281	58	0	23	0	161	6	1	2	25	0	5
SamsungC_T_SDG_2022	146	19	0	26	0	36	31	0	0	31	0	3
HanaFinancialGroup_ESG_2021	264	63	0	39	0	121	8	0	3	21	0	9
SKInnovation_ESG_2020	57	42	0	1	0	3	6	0	0	3	0	2
KakaoBank_SDG_2021	77	33	0	8	0	23	3	0	0	7	0	3
NAVER_ESG_2021	158	76	0	15	0	41	1	0	1	19	0	5
Kakao_ESG_2021	408	127	2	59	0	196	0	0	0	19	0	5
PoscoCorporate_ESG_2021	411	146	1	111	1	89	9	0	1	47	0	6
KT_ESG_2022	60	13	0	6	0	14	18	0	0	7	0	2
HyundaiMotorCompany_SDG_2022	235	69	0	33	0	65	31	0	3	27	0	7
IBK_SDG_2022	103	48	0	11	0	23	5	0	2	10	0	4
NCSOFT_ESG_2021	232	109	0	80	0	29	4	0	0	3	0	7
ShinhanFinancialGroup_ESG_2021	166	41	0	46	0	49	3	0	3	21	0	3
LGEnergySolution_ESG_2021	86	14	0	30	1	23	2	0	0	15	0	1
LGInnotek_SDG_2021	114	38	0	10	1	26	12	0	1	21	0	5
SamsungBiologics_ESG_2022	150	54	0	17	0	25	28	0	2	19	0	5
PoscoChemical_SDG_2021	196	86	0	17	0	61	8	1	0	22	0	1
LotteChemical_ESG_2021	136	53	0	18	1	45	9	0	0	8	0	2
SKIncorporated_ESG_2022	280	92	1	43	1	63	21	1	3	51	0	4



SKBioscience_ESG_2022	173	86	0	20	1	45	3	0	0	13	0	5
KIA_SDG_2022	117	34	0	11	1	23	17	0	2	24	0	5
HanwhaSolutions_SDG_2022	236	69	0	33	0	89	18	0	2	23	0	2
KoreanAir_ESG_2022	132	22	0	21	0	37	8	0	2	31	0	11
LGCorporation_ESG_2021	144	36	0	19	0	68	9	0	2	9	0	1
SKT_ESG_2021	277	86	0	24	0	78	30	2	2	46	0	9
LGH_H_ESG_2021	162	49	0	30	0	39	9	0	3	27	0	5
WooriBank_SDG_2021	121	33	0	22	1	36	3	0	2	19	0	5
HyundaiHeavyIndustries_ESG_2021	280	41	0	4	1	186	13	0	0	29	0	6
KEPCO_SDG_2021	155	65	0	7	0	22	7	1	1	48	0	4
LGElectronics_SDG_2021	99	32	0	12	0	21	4	0	1	20	0	9
KT_G_SDG_2021	326	112	0	93	0	55	20	0	17	20	0	9
DoosanEgerbility_SDG_2022	166	28	0	15	0	97	6	0	0	15	0	5
KBFinancialGroup_ESG_2021	198	50	0	65	0	43	11	0	4	22	0	3
HyundaiGlovis_SDG_2022	309	88	0	67	0	119	5	0	3	23	0	4
AmorePacificGroup_SDG_2021	163	50	0	34	0	29	10	0	2	32	0	6
KakaoPay_ESG_2021	408	127	2	59	0	196	0	0	0	19	0	5
SKHynix_SDG_2022	111	23	0	23	0	28	14	0	2	17	0	4
SamsungSDI_SDG_2021	157	37	0	14	0	87	6	0	0	11	0	2
SamsungElectronics_SDG_2022	234	107	0	42	1	50	1	0	1	23	0	9
HyundaiMobis_SDG_2022	255	58	0	21	0	115	7	1	2	44	0	7
SOIL_SDG_2021	216	74	0	14	1	84	8	0	2	32	0	1
SamsungFire_Marine_ESG_2022	119	43	0	13	1	41	8	0	1	7	0	5
HMM_ESG_2021	226	107	0	14	0	40	17	1	5	30	0	12
SamsungElectroMechanics_SDG_2021	238	74	0	66	1	42	6	0	2	35	0	12

## 제 V 장 결론 및 시사점

위의 연구결과에서 파악할 수 있듯이 국내 대기업들은 ESG 리포트 상에서 상대적으로 인식수준이 높은 조직문화, 규범준수 및 윤리, 다양성, 기술과 역량에 대한 영역 뿐만 아니라 인식수준이 상대적으로 낮은 조직의 건강, 안전 및 웰빙과 인력 가용성, 채용, 이동 및 이직률, 리더십, 생산성, 인력비용, 승계계획에 대한 영역에서도 그 공개 정도에 대해 진지한 고민이 필요해 보인다. 이처럼 기업들의 공개 정도가 낮은 이유는 이러한 데이터를 공개하였을 때 기업에 끼칠 위험에 대한 민감성 때문으로 판단할 수 있다. 하지만, 국내외의 다양한 이해관계자(stakeholder)들, 특히 재무적 투자자들로부터 투명한 ESG 데이터의 공개요구가 매우 커짐에 따라, 다른 두 영역인 환경(Environment)과 지배구조(Governance)의 공개정도에 비해 아직은 부족한 사회(Social) 분야의 인적자원영역에 대해서도 ISO 30414를 참조하여 그 공개범위를 확대해야 할 것이다.

서론에 언급한 것처럼 SEC는 구체적으로 공개해야 할 인적자원영역의 선정을 개별기업의 비즈니스 특성에 따라 자신에게 가장 적합한 것을 결정할 수 있도록 하고 있고, 글로벌 투자은행들도 인적자원영역에 있어서는 마찬가지로 영역의 정의나 계산방법들에 대해서는 구체적인 언급이 없다. 국내 K-ESG의 인적자원영역들도 그 영역이나 지표 개발에 대해서는 소극적인 것으로 보인다. ISO가 만들어진 이후 제안한 기준들이 다양한 영역에서 국제기준이 되는 시점에서, ISO 30414에서 제시하고 있는 인적자원영역과 지표를 조만간 범용적으로 대부분의 기업들이 사용하게 될 것으로 예상된다.

ISO 30414의 영역들을 국내 주요기업들과 비교할 때, 좀 더 상관성이 높은 결과를 얻기 위해 다른 기준이나 빅데이터 방식을 사용해 볼 필요가 있어 보인다. 그리고, 국내 주요기업들과 해당 기업들의 글로벌 경쟁사들과의 ESG

보고서상의 인적자원영역에 대한 비교를 통해, 산업별로 인적자본 정보공개 수준을 비교해 보는 것도 의미가 있어 보인다. 더불어, 본 연구의 보완적인 미래 연구로서 국내의 각 그룹사 및 산업별 ISO 30414 인적자원영역 언급 횟수의 차이를 살펴보고 그룹사 별 기업 문화 및 인적자원 관리능력과의 상관관계를 파악하는 연구로의 확장 가능성을 기대할 수 있을 것으로 사료되며 다양한 ISO 가이드라인을 변수화 하여 ESG 정보공개의 척도를 파악하는 연구로 발전시킬 수 있을 것이다.

## 참고문헌

- 산업통상자원부, 한국생산성본부 (2021), K-ESG 가이드라인 v1.0.
- 안태욱, 이희승, 이준서 (2021), “연관규칙 분석을 통한 ESG 우려사안 키워드 도출에 관한 연구,” *정보시스템연구*, 제30권 제1호, pp.123-149.
- 이재진(2020), 최초의 국제 표준 HR보고서 가이드 - ISO 30414 월간 인재경영 2020년 10월호, 제188호.
- 이재진(2021), HR 리더의 전략적 선택, ESG, 월간 인재경영 2021년 4월호, 제194호.
- 임효진 (2021), “텍스트 마이닝을 활용한 중소기업 ESG 이슈 분석,” *인문사회21*, 제12권 제4호, pp.469-481.
- 조대형 (2021), “ESG 글로벌 추진 현황과 사례 분석,” *인문사회21*, 제12권, 3호, PP. 2651-2662.
- 천상은, 박희태 (2021), “HRM 관점의 ESG 평가 영역 비교분석 연구,” *인적자원관리연구*, 제28권, 5호, PP.133-148.
- 한국기업지배구조원(KCGS) (2021), ESG 모범기준.
- Levenshtein, V. I. (1966), “Binary Codes Capable of Correcting Deletions, Insertions and Reversals,” *Soviet Physics Doklady*, Vol. 10, No.8, pp.707.
- Mihalcea, R., Tarau, P. (2004), “TextRank: Bringing Order into Text,” *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp.404-411.
- Trotman, A., Puurula, A., Burgess, B. (2014), “Improvements to BM25

and Language Models Examined,” ADCS '14: Proceedings of the 2014 Australasian Document Computing Symposium, pp.58–65.

<언론 및 웹사이트>

연합뉴스, 2021년 10월 20일자, ESG위원회 설치 편차…코스피 상장사 15% vs 자산 2조원대 55%.

인공지능신문, 2020년 5월 28일자, 리미니스트리트, 소프트웨어 유지보수 강화 위한 AI애플리케이션 출시.

<https://wiserloner.tistory.com/917> 검색.

[https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.TfidfVectorizer.html#sklearn.feature\\_extraction.text.TfidfVectorizer](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#sklearn.feature_extraction.text.TfidfVectorizer).

[https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.CountVectorizer.html#sklearn.feature\\_extraction.text.CountVectorizer](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html#sklearn.feature_extraction.text.CountVectorizer).