IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 March 2024, accepted 27 May 2024, date of publication 30 May 2024, date of current version 6 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407188

RESEARCH ARTICLE

Exploring Environmental, Social, and Governance (ESG) Discourse in News: An AI-Powered Investigation Through Knowledge Graph Analysis

SIMONE ANGIONI[1], SERGIO CONSOLI[2], DANILO DESSÌ[3], FRANCESCO OSBORNE[4,5], DIEGO REFORGIATO RECUPERO[1], AND ANGELO SALATINO[4]

[1]Department of Mathematics and Computer Science, University of Cagliari, 09124 Cagliari, Italy
[2]European Commission, Joint Research Centre (DG JRC), 21027 Ispra, Italy
[3]Knowledge Technologies for Social Sciences Department, GESIS Leibniz Institute for the Social Sciences, 50667 Cologne, Germany
[4]Knowledge Media Institute, The Open University, MK7 6AA Milton Keynes, U.K.
[5]Department of Business and Law, University of Milano Bicocca, 20100 Milan, Italy

Corresponding author: Diego Reforgiato Recupero (diego.reforgiato@unica.it)

ABSTRACT In recent years, the significance of Environmental, Social, and Governance criteria in assessing financial investments has grown significantly. This paper presents an AI-driven analysis of ESG concepts and their evolution from 1980 to 2022, with a specific focus on media sources from the United States and the United Kingdom. The primary data source utilized is the Dow Jones News Article dataset, providing a comprehensive and high-quality collection of news articles. The study introduces a novel technique for information extraction from news articles, involving the structuring of extracted data into a knowledge graph. The findings identified key trends associated with ESG aspects emerging in recent years. In the environmental dimension, we identified a pronounced emphasis on *climate change*, *renewable energy sources*, and *biodiversity conservation*. Within the social aspect, the analysis pointed out the increasing significance of issues such as *racism*, *gender identity*, and *human rights*, as well as the increasing role of *charities*, and the ethical challenges of modern *supply chains*. Finally, in the governance domain, the findings emphasized issues related to *corporate governance accountability*, *workplace ethics*, and the *conduct and remuneration of executives*.

INDEX TERMS ESG, knowledge graph, monitoring tool, extraction pipeline.

I. INTRODUCTION

In the last few years, Environmental, Social, and Governance (ESG) criteria become increasingly crucial to evaluate financial investments.[1] The European Parliament has acknowledged the significance of ESG ratings in its legislative efforts aimed at promoting an economy that genuinely serves the interests of the people. This led to several concrete initiatives, such as the introduction of the *EU taxonomy for sustainable activities*,[2] a resource seeks to establish a set of ESG standards for an organization's behavior, serving as a valuable tool for socially conscious investors when evaluating potential investments. The *environmental* aspect evaluates an organization's environmental impact and how it manages its use of natural resources, energy efficiency, waste management, and overall commitment to sustainability. The *social* criteria considers issues such as diversity and

The associate editor coordinating the review of this manuscript and approving it for publication was Mansoor Ahmed.

[1]Environmental, Social and Governance (ESG) rating activities - https://www.europarl.europa.eu/legislative-train/theme-an-economy-that-works-for-people/file-esg-rating.

[2]EU taxonomy for sustainable activities - https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities_en.

inclusion, labor practices, human rights, and community engagement. Finally, the *governance* aspect focuses on the internal processes and structures that guide an organization. This includes aspects such as corporate governance, business ethics, transparency, and the quality of leadership.

Even beyond the confines of the corporate sector, it is essential to acknowledge that ESG principles extend their relevance, acquiring universal importance in the broader ambition of promoting a sustainable and equitable global environment. Consequently, it is essential to monitor and analyse the portrayal and evolution of ESG-related concepts within the information landscape [1]. This examination would allow us to assess the changing perceptions of both media and public opinion on crucial issues such as sustainability and diversity [2]. Nevertheless, this undertaking is particularly challenging given the nuanced nature of these concepts and the difficulty in analysing them on a large scale.

Today, there are numerous news monitoring tools available that can be used to perform different kinds of analysis on the news (e.g., Brandwatch,[3] Brand24,[4] Repustate,[5] Cision Communication Cloud,[6] SentiOne,[7] and Meltwater[8]). However, current systems lack a sufficient representation of the nuanced dynamics of discourse, thereby making them incapable of supporting advanced queries related to the entities mentioned in news articles. For instance, while they can identify specific tags or keywords, they cannot extract the connections between them or the specific statements in which they are used. This limitation impedes their ability to perform a comprehensive analysis of the discourse about ESG.

To overcome this limitation, researchers have suggested various approaches to develop structured, interconnected, and machine-readable data frameworks for analysing news [3], [4]. Several of these representations employ semantic technologies, such as knowledge graphs. Knowledge graphs (KGs) are networks that consist of entities and their relationships, providing information in a machine-readable and understandable format within a specific domain [5]. In recent years, KGs have been increasingly acknowledged for their ability to organize structured data in a semantically meaningful way, providing effective support to a variety of AI systems in a variety of domains, such as medicine, research, education, robotics, manufacturing, social media, and many others [6]. Prominent instances of knowledge graph include DBpedia[9] [7], Google Knowledge Graph,[10] BabelNet,[11] and YAGO.[12] As discussed in a recent survey by Opdahl et al. [3], the creation of a KG from the news poses several challenges, such as performing quality named entity recognition (NER) and relationship extraction, the management of temporal information, entity linking, news source reliability, and different kind of co-reference resolution (pronoun, event, etc.). Large-scale knowledge graphs are frequently produced through a semi-automated process that integrates both structured and unstructured data. When the source data includes a large amount of text, these approaches typically use various natural language processing techniques for generating triples reflecting the key domain concepts and link them both to the original sources and to other information that allows users to assess their reliability [8]. Similar solutions were developed to characterize a variety of domains, such as research articles [9], medical information [10], tourism [11], educational resources [12], and social media [13].

This paper presents an AI-driven analysis of ESG concepts and their evolution from 1980 to 2022, focusing on media sources from the United States and the United Kingdom, including prominent publications such as The Guardian, The New York Times, and The Times. The primary data source for this analysis is the Dow Jones News Article dataset,[13] which offers a comprehensive and high-quality collection of news articles.

To facilitate this analysis, the study introduces an innovative technique for information extraction from news articles, which involves structuring the extracted data into a KG. This method employs advanced information extraction methodologies to distill relevant information from articles into structured statements represented as triples in the format `<subject, predicate, object>`. The operational pipeline developed for this process is generalizable and can be implemented on a conventional server, thereby avoiding the necessity for substantial computational resources, a requirement characteristic of current large-scale language models for processing massive amounts of data. The primary advantage of this innovative approach lies in its ability to facilitate the analysis of various entity types (e.g., organizations, persons, topics) while also establishing meaningful relationships between the entities based on predicates extracted from the articles. Consequently, it serves as an effective instrument for analysing a large volume of news content, deriving insights about key concepts, and understanding the development and changes in the discourse over time.

The resulting knowledge graph was employed to analyze the three core components of Environmental, Social, and Governance (ESG) and to identify the principal subjects emerging in recent years. In the environmental dimension, the analysis highlighted a pronounced emphasis on *climate change*, *renewable energy sources*, and *biodiversity conservation*. Within the social aspect, the analysis pointed out the increasing significance of issues such as *racism*, *gender identity*, *human rights* as well all the increasing role of *charities*,

[3]Brandwatch - https://www.brandwatch.com/.
[4]Brand24 - https://brand24.com/.
[5]Repustate - https://www.repustate.com/.
[6]Cision Communication Cloud - https://www.cision.com/.
[7]SentiOne - https://sentione.com/.
[8]Meltwater - https://www.meltwater.com/.
[9]DBpedia - https://www.dbpedia.org/.
[10]Google Knowledge Graph - https://developers.google.com/knowledge-graph.
[11]BabelNet - https://babelnet.org/.
[12]YAGO - https://yago-knowledge.org/.
[13]Dow Jones News Article dataset - https://developer.dowjones.com/datasets/details/news.

and the ethical challenges of modern *supply chains*. Finally, in the governance domain, the findings emphasized issues related to *corporate governance accountability*, *workplace ethics*, and the *conduct and remuneration of executives*.

More in detail, the contributions of our paper are the following:

- We present an AI-driven analysis of the news discourse around ESG concepts from 1980 to 2022;
- We propose a general and automatic pipeline to create a KG from a set of news documents;
- We demonstrate how to produce several analytics regarding entities and statements from a KG extracted from the news;
- We report an evaluation of the information extraction pipeline, showing excellent accuracy.

The remainder of this paper is organized as follows. Section II discusses related works about KGs on news and various methodologies to create them. Section III describes the general pipeline for KG generation and reports its evaluation. Section IV details the data source and offers an overview of the resulting KG centered on ESG aspects. Section V discusses the results of the analysis. Finally, Section VI ends the paper with conclusions and future works where we are headed.

II. RELATED WORK

In this section, we review the current state of the art regarding news monitoring approaches (Section II-A) and knowledge graphs on news (Section II-B).

A. NEWS MONITORING APPROACHES

News analysis and monitoring is a broad research area that encompasses various tasks, including first story detection, clustering, trends detection, event detection, question/answering, summarization, and fake news detection [14], [15], [16], [17], [18].

Traditionally, these tasks involved tracking a set of keywords, mainly extracted applying Term Frequency Inverse Document Frequency (TF-IDF) [19], and were primarily used for conducting small-scale analyses [20], [21]. Such methods required significant manual effort from media scholars and practitioners, restricting the scale of the analyses. Furthermore, the absence of semantic matching in these approaches limited their ability to fully represent the diversity and complexity of the news media landscape. This limitation became particularly evident with the emergence of new forms and sources of news, including social media, blogs, podcasts, and citizen journalism [22].

For instance, Lloyd et al. [23] developed an approach for analysing news media [23] and blog posts [24] to perform juxtaposition, temporal, and spatial analyses. The most pressing challenge was entity disambiguation [25], which they solved by combining the syntactic similarity of their surface form and the co-occurrence analysis with other entities. Nowadays, such a challenge can be easily tackled with the support of notable knowledge graphs like DBpedia [7] or Wikidata [26]. Indeed, Piskorski et al. [27] mapped entities in news from multiple languages relying on BabelNet [28]. Similarly, Scharl and Weichselbraun [29] conducted a study on the media coverage of U.S. presidential elections, and in this limited analysis the authors extracted entities by matching terms in a keyword list.

Tanev et al. [30] proposed a method for global crisis monitoring that extracts information about violent and disaster event. Their pipeline consists of news geo-tagging, automatic pattern learning, pattern specification language, information aggregation. Specifically, they identify locations by mapping n-grams to a multi-lingual gazetteer whereas they extract events or actions with pattern-matching rules. However, such an approach fails to capture the subtle linguistic features that are common within news articles.

More advanced solutions employ machine learning algorithms. For instance, Téllez et al. [31] developed an approach for extracting information from natural disaster news reports. It first identifies candidate text segments, then it employs a support vector machine (SVM) and a Naïve Bayes to classify such statements according to five different types of disasters. However, this solution lacks generalisability. Indeed, it is limited to a low number of categories that can be classified, and it requires the identification of the most suitable set of features when applied to a new domain.

Recently, this field went through a paradigm shift benefiting from a number of innovative solutions being developed in the field of artificial intelligence [32]) (e.g., large language models [33]) and semantics technologies (e.g., KGs [34], [35]). For instance, large language models have been employed to summarize news [36]. Deep learning has been employed to identify text chunks (e.g., entities) within news that are worth analysing [37], as well as fake news [38]. Recently, knowledge graphs have been used to structure the news content in a machine-readable format and further support the aforementioned tasks [3], [4]. To the best of our knowledge, there has not been any work employing a KG-driven analysis of the ESG space.

B. KGs ON NEWS

The primary objective of employing KGs in news analysis is to represent and establish relationships between various entities in the news domain, such as people, places, events, topics, and facts. This systematic overview can lead to a more insightful analysis of the changes in discourse over time. For instance, Al-Obeidat et al. [39] constructed a KG that represents news related to COVID-19. This KG offers a platform for researchers, data analysts, and data scientists from various sectors to explore and suggest solutions for the challenges that COVID-19 creates for the global society. Tan et al. [4] focused on electronics and supply-chain industry news to build a KG on causal relations. This KG can be utilized by companies to make informed decisions and predictions. Liu et al. [40] proposed a KG-based news recommendation system. The distinctive feature of their

KG is that it records the topic context of the news, links entities with collaborative edges derived from the users' clicks and the co-occurrence in the news articles, and removes news-irrelevant relations. Rospocher et al. [41] developed an event-centric knowledge graph based on news sources. Their methodology emphasizes representing a temporal dimension that captures the long-term development and histories of all entities involved. Fu et al. [42] developed a multi-domain KG to support fake news detection. Their tool employs the knowledge graph to produce background information, semantically link news articles, and improve the learning and classification of news content. In this context, the knowledge graph enhances their methodology's ability to generalize effectively across single, mixed, and multiple domains, surpassing existing state-of-the-art techniques in multi-domain fake news detection. Opdahl et al. [3] surveyed research methods and approaches that employ semantic KGs for producing, distributing, and consuming news.

In this paper, we generate a KG as an intermediary phase in our AI-powered analysis of the ESG sectors. Unlike previous studies, our KG is designed primarily to facilitate the analysis of how the ESG discourse evolved over time, through a detailed depiction of various types of entities.

III. AUTOMATIC GENERATION OF ESG-KG

In this section, we will discuss the generation of the KG from a repository of news. The pipeline we developed, illustrated in Figure 1, consists of two primary stages. Initially, it employs a *Text Parsing Module* to extract entities and their relationships from a collection of news articles. The knowledge graph is generated through a three-step process in the subsequent stage. The *Entity Extraction Module* identifies key entities and categorizes them by type. Following this, the *Relationship Extraction Module* extracts the relationships between these entities from the news articles. Finally, the *Triple Refinement Module* finalizes and refines the resulting triples to produce the completed knowledge graph. In the following subsections, we will first describe the Text Parsing Module (Section III-A) and then outline the KG generation stage (Section III-B).

A. TEXT PARSING MODULE

The Text Parsing Module is based on the Stanford CoreNLP[14] suite, a comprehensive suite of natural language processing tools developed in Java. It is a robust and thoroughly tested set of tools, popular across academic, industrial, and governmental circles. This tool processes raw text in natural language through a combination of rule-based techniques, probabilistic machine learning, and deep learning to conduct various analyses. It determines the root forms of words and their grammatical functions, identifies named entities (e.g., companies, individuals, locations) and extracts dates, times, and numerical values. Furthermore, it annotates sentence structures by identifying syntactic phrases or dependencies and recognizes noun phrases that reference the same entities.

More specifically, this module uses the Part-of-Speech Tagger (PoS Tagger) to assign tags to each word in the given text. Figure 2 shows the result of this process for a sentence from the ESG news corpus. The PoS tagger identifies and classifies each token based on its grammatical category (e.g. preposition (PRP), verb (VB), noun (NN), adjective (JJ), etc). In the sample sentence, the PoS Tagger identified *E.P.A* as a proper noun (NNP), *responsibility* as a singular noun (NN), and *options* as a plural noun (NNS). Moreover, it correctly identified the verbs *think*, *has*, and *give*.

The module also builds a dependency tree of each sentence (see Figure 3). This is a representation of grammatical structure that maps out the relationships between tokens in a sentence. A dependency tree includes:

- **Nodes and Relationships**. A dependency tree is built on three main components: i) nodes, every word in the sentence is a node in the tree, ii) edges, representing the grammatical relationships between these words, and iii) root, representing usually one word (often the main verb) serving as the root of the tree, from which branches extend to other parts of the sentence.
- **Types of Relationships**. Dependency trees capture various types of grammatical relationships, such as subject, object, modifier, etc. For example, in the sentence in Figure 3 the dependency tree captures *E.P.A* as the subject of the verb *has* with object *responsibility*.
- **Directional Relationships**. The edges in the tree are directional, indicating which word is the ''head'' (the word that provides grammatical structure) and which is the ''dependent'' (the word that depends on the head for its grammatical role).

In summary, dependency trees provide a clear and structured way to represent the grammatical relationships between words in a sentence, enabling automatic usage of the natural language for complex NLP applications.

B. KNOWLEDGE GRAPH GENERATION MODULE

This section illustrates the pipeline for extracting entities and relationships from news articles and generating triples for the ESG-KG. The approach, shown in Figure 1 is organized in three steps: i) Entity extraction, ii) Relationship extraction, and iii) Triple extraction. In the following, we will analyze each step of the pipeline.

1) ENTITY EXTRACTION SUBMODULE

This module detects nominal phrases which will be used as entities for the KG. Nominal phrases are word groups with a noun or pronoun as its main word, along with any modifiers, determiners, and complements that provide additional information about the noun (e.g., 'long news article'). To detect nominal phrases, we start from the dependency tree extracted by the previous module. In the dependency tree, each token (or word), is associated with its part of speech (POS) tag which clarify if that token is

[14]Stanford CoreNLP - https://stanfordnlp.github.io/CoreNLP/.

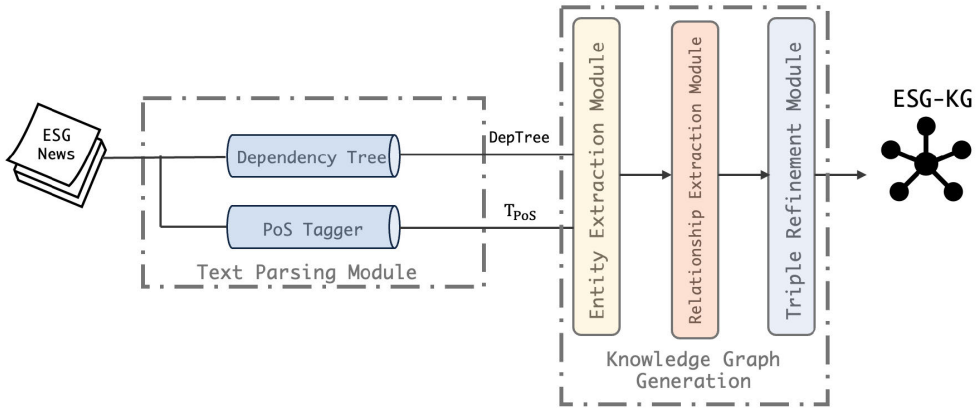


FIGURE 1. Knowledge graph generation pipeline.

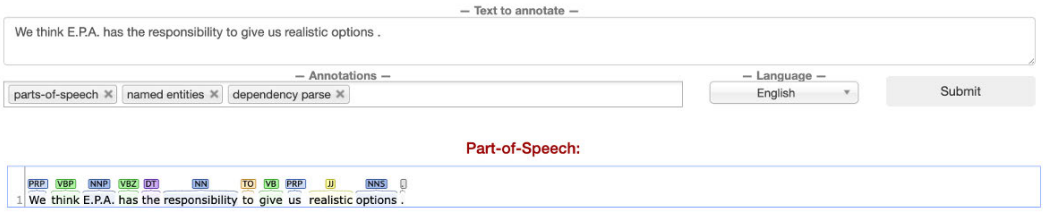


FIGURE 2. Part of speech tags.

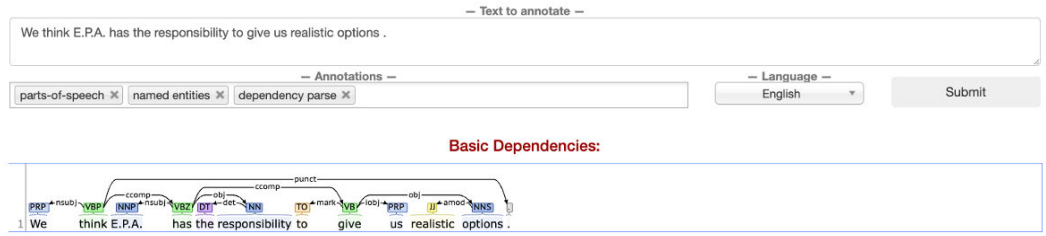


FIGURE 3. Dependency tree.

a noun (NN), verb (VB), adjective (JJ), personal pronoun (PP), and so on. To minimize irrelevant entities, our approach involves two steps. First, we consider only the phrase that contains at least one token classified as a noun (NN). Second, we expand from this noun token through the dependency tree to incorporate additional tokens, guided by the grammatical relationships specified in Table 1. The output of this module is a set of entities E_s associated with each sentence s from the news corpus. For example, starting from the noun (NN) 'panel', we can consider the adjective (AMOD) 'solar' to identify the entity 'solar panel'.

2) RELATIONSHIP EXTRACTION SUBMODULE

This module detects the relationship between entities. For each sentence s all the shortest paths of the dependency tree between each pair of entities $(e_i, e_j)|e_i, e_j \in E_s$ containing a verb are selected. This process will yield several types of paths between entities, some of which may be more reliable to derive a relationship depending on the source data and the style of writing. It is thus advisable to analyse the paths and determine the most suitable ones for identifying relationships in the particular context [43].

We also took this approach for producing the ESG KG. When this process was applied to the dataset of news articles, it yielded approximately 15K paths. From these paths, we computed the most frequent patterns i.e., the ordered list of dependencies that compose a path (e.g. [*nsubj*, *obj*]).

TABLE 1. Grammatical relations used to extract entities.

Relation	Usage
nmod	used for nominal modifiers of nouns or clausal predicates
nummod	used to link a noun to a numeric modifier
amod	used to link an adjective modifier to a noun
conj	a conjunct is the relation between two tokens connected by a coordinating conjunction, such as *and*, *or*, etc.

We ordered these patterns from most to least frequent based on their occurrence in the corpus. Next, the top 20 frequent patterns, with frequencies ranging from 79 to 1098, were manually reviewed by three researchers working in the NLP field. Their task was to assess a random sample of 20 triples for each pattern as *valid* or *not valid*. Each evaluator was tasked to assess the correctness of all 400 relevant triples. More precisely, to be annotated as valid, a triple should capture the semantics of the portion of the sentence where it was extracted. For example, the triple `<Mr. Lewis, give, quixotic guided tour>` extracted from the sentence *'Mr. Lewis gives the reader a quixotic guided tour through Silicon Valley while showing how its success stories revolutionized American business.'* with path [*nsubj*, *obj*] was considered valid by the annotators. On the other hand, the triple `<air, rising, hot day>` from the sentence *'Howe says it was discovered by cows drawn to cool air rising from the mouth of the cave on a hot day.'*, with path [*acl*, *obj*] was discarded as not valid by the annotators.

A majority vote was used to label each triple as correct/incorrect and only the subset of patterns with a prevalence of correct triples (i.e., more than 10) were considered reliable and kept in the result list. The set of valid patterns is referred to as P_{valid}.

Finally, for each sentence s_i all the shortest paths containing a verb v_i of the dependency tree between each pair of entities (e_m, e_n) were computed. The resulting set of paths was filtered, selecting only those that match a pattern in P_{valid}. The entities (e_m, e_n) and verb v are used to create triples T in the shape $< e_m, v, e_n >$.

3) TRIPLE EXTRACTION SUBMODULE

This module performs three main tasks: 1) relation refinement, 2) entity refinement, and 3) triple refinement.

a: RELATION REFINEMENT

The set of triples T, generated in the previous step, may contain triples with similar meanings but using different verbs, e.g., `<company, build, 200 unit motel>`, `<company, construct, buildings>`, `<craftsmen, create, accommodation>`. This step aims to find the best predicate label r for each relation verb v in a triple $< e_m, v, e_n >$ and to map v to r in the resulting triple. In this phase, it is recommended to reduce the space of possible relationships by analysing the resulting verbs and clustering them in a smaller set of well-defined relationships [43]. We thus need to find the verbs that have a similar meaning. For this purpose, we used Wordnet.[15] WordNet is a large lexical database of English. This knowledge base groups nouns, verbs, adjectives, and adverbs into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked using conceptual-semantic and lexical relations. Consequently, we associated each verb with the synonyms using the synset from WordNet classes. If the synset similarity among two verbs, v_1 and v_2, was higher than the threshold (0.7) they were inserted in the same cluster. Finally, for each relation verb v in the dataset, we replace it with the predicate label r consisting of the lemma of the most frequent relation in the cluster of v. Otherwise, we map it to itself if v was an outlier and not clustered.

This method was applied to the 393 verbs found in all the triples extracted from the ESG news dataset. It produced a final set of 57 predicates.

b: ENTITY REFINEMENT

Nominal phrases may refer to the same entity in different ways; for instance, *President Obama*, *B. Obama*, *Barack Obama* likely denote the same person. To address this issue and identify nominal phrases that refer to the same entity, this module utilizes a sentence transformer model.

Given the set of all entities E, the module creates an index based on the tokens contained by the entities. The index links each token to all the entities that include it. For example, in the index, the token *Obama* is linked to all the entities which include it, such as *Barack Obama*, *President Obama*, *former president Barack Obama*, *Barack Obama's Administration*, *Michelle Obama*, and so on. We then compare two entities $e_i, e_j \in E$ if they share at least one token. The comparison is performed by using the state-of-the-art framework SentenceTransformers[16] and encoding the entities with the *all-mpnet-base-v2*[17] transformer model. We chose this model since it showed state-of-the-art performances on a multitude of tasks including semantic text similarity.[18] Entities with a cosine similarity equal to or greater than a threshold $eth_{merge} = 0.9$ (empirically calculated) are merged. For example, if the entity e_i and e_j have a cosine similarity greater than 0.9, then e_i and e_j are inserted into the same cluster.

Finally, for each entity e in the dataset, we replace it with the entity label r consisting of the lemma of the most frequent entity in its cluster. Otherwise, we map it to itself if e was an outlier and not clustered.

c: TRIPLE REFINEMENT

Similarly to the entity refinement step, we used a sentence transformer model to detect and merge triples with the same meaning. Given the set of all triples, let us say T, the module creates an index based on the tokens contained in the subject and object of the triples. The index links each token to all the triples that include it. Then, it compares two triples $t_i, t_j \in T$ if they share at least one token. The comparison is performed by using the state-of-the-art framework *SentenceTransformers* and encoding the triples with the *all-mpnet-base-v2* transformer model. Triples that

[15] Wordnet - https://wordnet.princeton.edu/.

[16] SentenceTransformers - https://huggingface.co/sentence-transformers.

[17] all-mpnet-base-v2 - https://huggingface.co/sentence-transformers/all-mpnet-base-v2.

[18] SentenceBERT - https://www.sbert.net/docs/pretrained_models.html.

have a cosine similarity equal to or greater than a threshold $th_{merge} = 0.9$ (empirically calculated) are clustered together.

As final step, the resulting triples are linked to the original papers and used to construct the knowledge graph. Each triple is also associated with its *support*, i.e., the number of news articles from which it was extracted. This *support* score can serve as a criterion for evaluating the reliability of a triple. A triple with high support indicates frequent occurrence across diverse news sources, suggesting it is a recognized element of public discourse. Conversely, a triple with low support, appearing only sporadically, might be considered less reliable.

C. EVALUATION

We evaluated our pipeline by assessing its accuracy on a sample of triples. First, we selected a random sample of 200 statements, equally distributed among high-support (support greater than 10), medium-support (support lower or equal to 10 and greater than 5), and low-support triple groups. Each triple underwent evaluation by three reviewers. The evaluators were given both the triple and the original sentences from which the triple was derived. They had to mark the triple as it as 1 if it correctly reflected the content of the news articles, and 0 if it did not. The average agreement between the annotators was 0.89, indicating they mostly agreed with each other.

We evaluated the 200 statements produced by the pipeline against the majority vote of the three annotators, yielding an accuracy of 0.85. Individual rater estimates ranged from 0.85 to 0.93. This indicates that the pipeline can extract triples with good accuracy.

IV. THE ESG KNOWLEDGE GRAPH

In this section, we will outline the data source and offer an overview of the resulting KG.

A. DATA SOURCE

The Dow Jones News Datasets is a vast collection of 15, 105, 283 news articles in various languages. The dataset offers 13 English sources, including notable ones like The Wall Street Journal, New York Times, and The Guardian, contributing to a total of 7.3 million distinct news items. Each entry in the Dow Jones Dataset includes: i) the full text of the article, ii) its title, and iii) a set of relevant metadata.

The metadata can be broadly classified into two categories: i) general metadata, and ii) article content metadata. The general metadata contains information about the news source (e.g., source name, source language, publisher name, publisher region), time-related details (e.g., publication date, ingestion date), and the unique identifier. The article content metadata contain various types of information extracted or derived from the full news article. The main ones are:

- Subject: contains information about the overall topic discussed in the news;
- Region: contains all the countries or regions mentioned in the news;
- Company: this metadata articulates in three different fields: 1) the name (or code) of the companies with high relevance to the news (e.g., when a company is the primary subject of the article), 2) companies that are mentioned in the text of the news with low relevance to the news, and 3) a field that includes companies that are related to the ones mentioned in the text although they do not appear in the news. The relation can be both of a legal nature or other types of relations (e.g., collaboration).
- Market Index: contains the information about the market index relevant to the news.

To select a repository of news articles about ESG we considered all news from 1980 to 2022 including keywords about Environmental, Social, and Governance either in the text body or in the metadata fields. The keywords were selected by sector experts analyzing and filtering common terms and phrases about Environmental, Social, and Governance resulting in a vocabulary of 454 different keywords: 355 keywords for Environmental, 53 for Social, and 46 for Governance. The final collection consists of approximately 850, 000 news articles: 500, 000 on environmental topics, 290, 000 on social issues, and 60, 000 on governance.

B. ESG-KG

We applied the pipeline described in Section III to the set of 850, 000 ESG news articles. The resulting KG includes over 7.2M statements and 4M entities. To structure the statements, we employed a lightweight ontology, since the main purpose of the Knowledge Graph is only to aid in analyzing the news. The ontology defines four main Classes: i) aggregated statement, ii) fine-grained statement, iii) News, and iv) Entity. The ontology also specifies 57 object properties derived from the predicates defined in Section III-B3.

The ontology also maps the statements using the original verb with their version using the 57 predicate obtained by clustering them.

Each statement in ESG-KG includes:

- *rdf:subject*, *rdf:predicate*, and *rdf:object*, which provide the reification of triples within a rdf:Statement;
- *provo:wasDerivedFrom* which provides provenance information and lists the DNA-IDs of the news from which the statement is derived;
- *esg-kg:statement_negated* which is a boolean whose value depends on the form of the statement in the news text. It will be True if the statement was derived from a negative sentence, False otherwise;
- *esg-kg:original_triple* which lists the fine-grained versions of the statement.

Additionally, each news ID is linked to *xsd:date* that provides the publishing date of the news, and *esg-kg:source* that provides the journal source name where the news was originally published.

V. ANALYSIS OF THE ESG DISCOURSE IN THE NEWS

In this section, we present and analyze the findings of our study on Environmental, Social, and Governance (ESG)

TABLE 2. Type of entities and their frequency. GPE stands for Geopolitical Entities, whereas NORP represents nationalities, religious, and political groups.

Entity Type	# Entities
PERSON	519K
ORG	359K
CARDINAL	244K
GPE	232K
NORP	116K
DATE	101K
LOC	32K
ORDINAL	30K
PERCENT	19K
TIME	16K
MONEY	7.2K
EVENT	2K
OTHER	2.2M

discourse in news media. The analysis is based on various analytics produced from the knowledge graph.

A. DIACHRONIC ANALYSIS

Figure 4 shows the distribution of the three main topics (Environmental, Social, and Governance) across time. Historically, environmental issues have dominated, accounting for 55% to 75% of coverage. However, there is a noticeable trend of increasing focus on social issues, which has grown from approximately 20% in 1980 to nearly 40% by 2022. This shift appears to be driven by a heightened interest in subjects like ethics, racism, gender identity, and global human rights.

B. ENTITY AND STATEMENT DISTRIBUTION

The three key pillars of ESG encompass a broad range of entities: the environmental component includes 2M entities, the social aspect covers 361K entities, and the governance section comprises 209K entities. The entities are typed according to the Named Entity Recognition (NER) tool provided by Spacy. The entity types and their frequency are reported in Table 2.

In Figure 5 displays how entities are distributed based on the number of statements they are associated with. Likewise, Figure 6 reports the distribution of all statements over the number of articles from which they were extracted. The KG covers a total of 3.8M statements: 3M statements related to environmental topics, 600K for social issues, and 236K concerning governance. Both entities and statements exhibit a typical long-tail distribution, where a small number of key entities and statements occur frequently, and a vast majority of less significant ones appear very infrequently.

C. ESG ANALYSIS

In this section, we will perform an in-depth analysis of each of the three pillars of ESG - Environmental, Social, and Governance. First, we report the top ten religious or political groups (Table 3), geopolitical entities (Table 4), notable persons (Table 5), and organizations (Table 6) for each of them.

A key observation is the prominent position of the USA in ESG discourse, evidenced by the top three groups being Democrats, Republicans, and Americans. Furthermore, the most frequently mentioned individuals are US Presidents, including Bush, Obama, Clinton, and Trump. The United States is also the country most frequently cited in articles related to the Environmental and Social aspects, highlighting their perceived leadership status on these topics. In contrast, China takes the lead in discussions on Governance. Beijing and Japan are often mentioned in discussions about employee rights, ranking them fourth and fifth, respectively, among the most mentioned countries for Governance.

In the following, we delve into each of the three ESG domains, examining the evolving trends of key entities within each area. To do so, we have computed the number of times each entity appears in each year. Following this, we applied linear regression to the yearly distributions of these entities, with the regression line's slope serving as an indicator of the trend's trajectory. A more pronounced slope denotes a more rapid escalation in media coverage for the specified entity. This is a technique commonly applied to detect key trends, e.g., in research topics [44]. We report the results in Table 7-9, where *slope_10* indicates the trend in the last 10 years and the *slope_5* indicates the trend in the last 5 years. In order to highlight common themes, we manually clustered the related entities and highlighted them in the same colour.

In the following, we report an analysis of the major trends of each of the three macro themes based on both these analytics and an inspection of the relevant statements and articles in the KG.

1) ENVIRONMENTAL

The environmental aspect of ESG focuses on a company's impact on the natural world and its management of environmental risks. Table 7 displays the entities that have demonstrated the most significant growth in mentions over the past ten and five years.

We analysed the network of entities and their statements and identified three main topics that have gained more prominence in recent years:

- **Climate Change** and **Carbon Emissions** (orange in Table 7): These entities are central to discussions about measuring carbon footprints, implementing initiatives to reduce greenhouse gas emissions, and formulating strategies to mitigate climate change effects [45]. The heightened visibility of these entities in news narratives highlights the escalating importance of adopting concrete measures to combat climate change.
- **Energy Efficiency** (blue): The positive trends indicate a growing emphasis on the use of renewable energy sources and the adoption of energy-saving practices in public discussions. This shift towards energy efficiency reflects a broader societal and economic recognition of the benefits associated with sustainable energy practices. As the urgency to address climate change intensifies, the

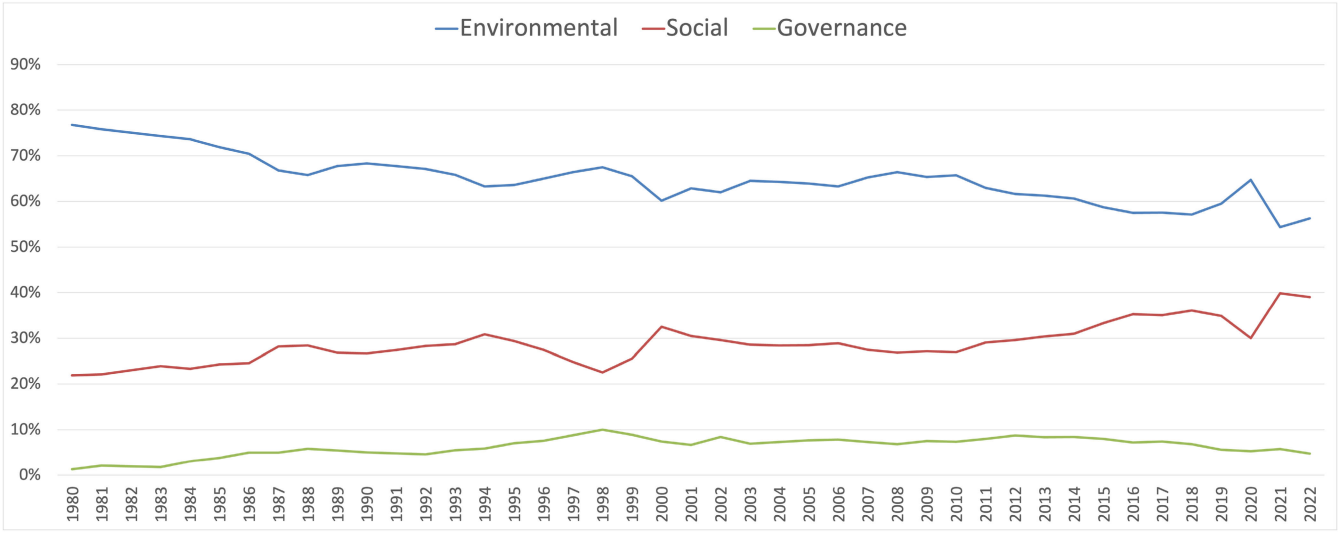


FIGURE 4. News distribution for year.

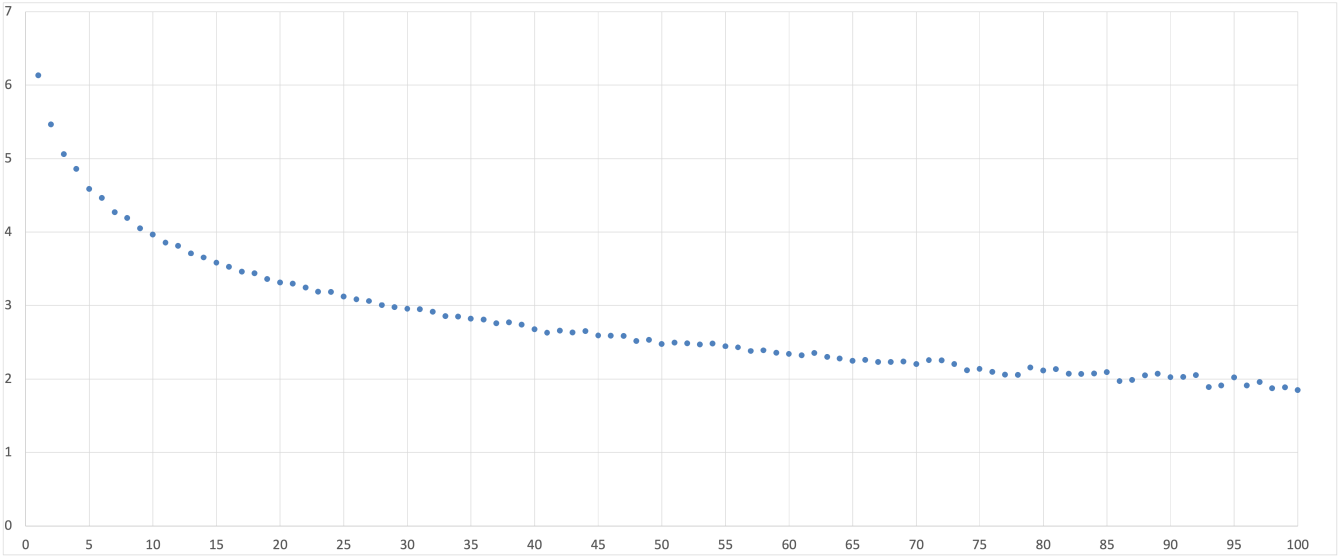


FIGURE 5. Entity distribution over the number of relevant statements (logarithmic scale).

TABLE 3. Top 10 NORP entities.

environmental		social		governance	
Democrats	7457	Democrats	1699	Democrats	434
Republicans	6818	Republicans	1435	Republicans	369
Americans	5111	Americans	1324	Americans	321
Russians	1273	Jews	460	European union	118
Democrat	901	Palestinians	405	Japanese corporation	86
Palestinians	894	Chinese government	376	European commission	65
Europeans	741	Russians	345	Chinese authority	48
Germans	495	Muslims	317	European community	35
British government	249	African	291	British company	25
Italians	214	Islamic State	247	German company	23

push for more efficient energy usage and the transition to renewables becomes a central theme in policy, industry, and community conversations [46].

- **Biodiversity** and **Land Use** (green): The upward trend over the past five years underscores the significance of these issues. Specifically, the focus on

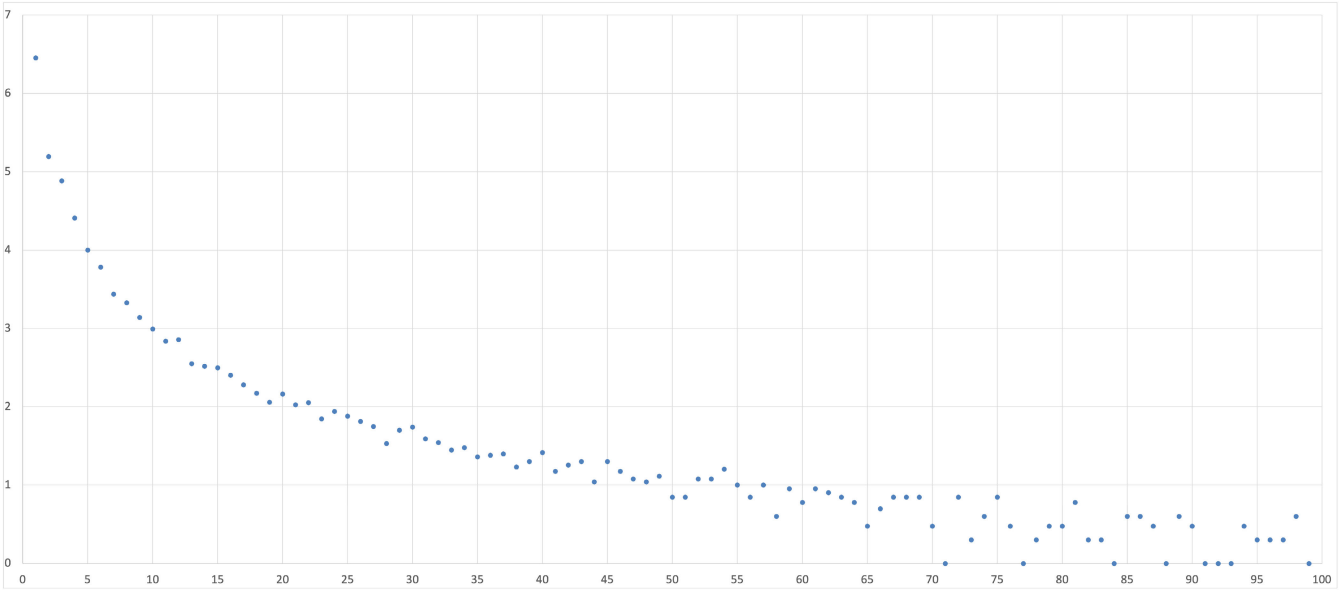


FIGURE 6. Statement distribution over the number of news in logarithmic scale.

TABLE 4. Top 10 geopolitical entities.

environmental		social		governance	
United States	13587	United States	5469	China	2078
China	7600	China	4310	United States	1355
California	5029	Russia	1822	Russia	466
Russia	4527	U.S.	1811	Beijing	423
U.S.	4358	Israel	1588	Japan	272
Washington	3877	Britain	1434	America	261
America	3308	Washington	1369	Britain	246
New York	3005	America	1166	India	214
Israel	2989	UK	932	UK	191
Japan	2729	France	712	Australia	146

TABLE 5. Top 10 person entities.

environmental		social		governance	
Bush	4727	Biden	1209	Trump	221
Obama	3478	Trump	1007	Bush	180
Clinton	2162	Bush	890	Obama	162
Johnson	1812	Johnson	688	Johnson	159
Brown	1709	Obama	667	Clinton	156
Trump	1705	Clinton	595	Brown	85
Mr. Reagan	1466	Brown	455	Greg Abbott	78
President Reagan	947	Putin	444	Mrs. Clinton	72
McCain	792	Jackson	421	Mr. Smith	70
Miller	704	Harry	306	Mr. Dimon	66

deforestation and biodiversity highlights the media's increasing concern and interest in the conservation of natural habitats, ecosystems, and biodiversity. This focus also aligns with global efforts to achieve biodiversity conservation targets and sustainable development goals [47], emphasizing the need for a holistic approach to environmental stewardship that includes protecting diverse ecosystems and ensuring responsible land use.

2) SOCIAL

In recent years, the social dimension of ESG has gained significant prominence, particularly concerning social issues such as racial justice, gender equality, fair labor practices, and economic inclusion. Table 8 presents the entities showing the most notable positive trends in this area. Similar to before, we can highlight several key topics:

- **Racism** (orange in Table 8): Over the last five years, the topic of racism has gained significant attention

TABLE 6. Top 10 organization entities.

environmental		social		governance	
Congress	12828	Congress	2807	Congress	1431
EPA	6491	Taliban	963	Microsoft	1212
White House	4243	White House	924	Apple	882
Senate	4062	Senate	849	Google	819
House	3585	United Nations	804	SEC	706
NASA	2649	Supreme Court	707	white house	417
Environmental Protection Agency	2242	House	661	Justice Department	396
Fed	2148	State Department	628	Intel	395
Ford	1810	EU	612	Facebook	390
Dow Jones industrial average	1765	Facebook	523	Fed	385

TABLE 7. Environmental entities.

Entity	Freq.	Slope 10 Years	Slope 5 Years
climate change	4818	58.01	99.20
temperature	4047	20.47	51
carbon emission	910	5.33	16.40
coal	1492	2.56	11
greenhouse gas emission	919	3.56	9.60
global warming	2831	1.76	7.20
natural gas	2508	-9.63	5.10
rising temperature	184	2	4.70
oil industry	585	-1.84	4.40
greenhouse gas	959	0.07	1.90
global temperature	145	0.38	0.70
air pollution	1073	4.57	0.30
solar panel	834	2.13	8.90
wind power	130	0.42	2.30
wind farm	199	0.62	1.90
solar farm	43	0.97	1.60
energy efficiency	444	-1.07	1.30
solar energy	251	0.22	1
solar power	253	-0.06	1
wind energy	45	-0.29	0.40
deforestation	223	1.68	4
fertiliser	22	0.94	2.30
biodiversity	249	3.03	2.20

TABLE 8. Social entities.

Entity	Freq.	Slope 10 Years	Slope 5 Years
racism	1858	46.72	118.10
racist	203	6.56	18.20
racial diversity	83	0.52	2.30
gender identity	81	2.58	7.10
gender equality	86	1.10	2
charity	413	10.87	31.20
community service	22	0.20	0.70
human rights	1370	13.08	41.30
responsibility	875	12.29	37.20
ethic	30	0.49	1.30
supply chain	46	1.76	4.30

within the social sphere, its rise fueled by several pivotal events related to this issue in recent years. This is in part due to high-profile incidents and movements that have sparked global conversations and protests. Events such as the Black Lives Matter movement, which gained renewed momentum following the death of George Floyd in the United States, have been central in bringing issues of racial injustice and police brutality to the forefront of media coverage.[19] Similarly, the discussions around systemic racism, racial profiling, and the need for reform in various institutions have been amplified by these incidents. There has been also an increased awareness and examination of colonial legacies, xenophobia, and racial disparities exacerbated by the COVID-19 pandemic [48]. These discussions have extended to various sectors, including education, healthcare, and the corporate world, where there is a growing call for diversity, equity, and inclusion initiatives [49].

[19]The global impact of George Floyd: How Black Lives Matter protests shaped movements around the world - https://www.cbsnews.com/news/george-floyd-black-lives-matter-impact/.

- **Gender identity** (blue): The topic of gender identity has also seen a significant rise in media coverage and public discourse, reflecting a broader shift towards recognizing and respecting diverse gender expressions and identities. This increased attention is part of a larger movement towards gender inclusivity and the rights of transgender and non-binary individuals [50]. High-profile legal battles over transgender rights, debates around gender-neutral bathrooms, and the inclusion of gender identity in anti-discrimination laws have been focal points in the media [51]. The corporate world is also responding to the growing recognition of gender diversity [52]. Many companies are implementing more inclusive HR policies, such as gender-neutral dress codes, the use of preferred pronouns, and support for employees undergoing gender transition. These changes are part of a broader effort to create inclusive work environments that respect and affirm diverse gender identities.
- **Charities and community service** (green): The media's focus on charities and community services has intensified, reflecting a heightened awareness and engagement with social issues at the grassroots level [53]. This trend is fueled by a growing recognition of the vital role these organizations play in addressing societal challenges, from poverty and homelessness to education and mental health. The response to crises like the COVID-19 pandemic has been significant, with coverage highlighting both the generosity of individuals and the crucial role of charitable organizations in providing

TABLE 9. Governance entities.

	Entity	Freq.	Slope 10 Years	Slope 5 Years
	board	1301	-0.58	5.70
	governance	40	0.72	2.10
	staff	317	3.75	8.90
	worker	614	2.34	5.40
	employee	1487	-1.05	1.70
	ethic	30	0.50	1.30
	equity	60	0.33	0.50
	payment	338	1.13	4.30
	executive compensation	119	0.22	0.90

relief and support [54], [55]. The rise of digital platforms has also transformed the landscape for charities and community services, making it easier for them to reach potential donors, volunteers, and those in need of assistance [56]. Social media campaigns, crowdfunding for social causes, and virtual volunteering opportunities have become increasingly prominent, facilitating greater engagement and support from the public [57].

- **Human Rights** (yellow): This increased focus is driven by global advocacy for a wide range of rights, including freedom of speech, the right to privacy, labor rights, and the protection of individuals against discrimination and abuse. Media coverage of human rights issues has been pivotal in bringing to light violations and conflicts around the world, from the plight of refugees and the treatment of ethnic minorities to the crackdown on freedom of expression in various countries [58]. In the corporate sphere, the emphasis on human rights has led to greater scrutiny of business practices, particularly in supply chains where labour rights violations or environmental degradation can impact local communities [59].
- **Supply Chain Management** (grey): This topic highlights growing concerns about sustainability, ethical sourcing, and labor practices [60], [61]. This rise in attention is in response to the complex global networks that define modern production and distribution systems, where issues in one part of the supply chain can have widespread implications [62]. Media coverage has played a crucial role in highlighting cases of unethical supply chain practices, such as child labor, unsafe working conditions, and environmental degradation.[20]

3) GOVERNANCE

ESG governance pertains to the internal practices and policies guiding a company's management and decision-making processes. Unfortunately, these nuanced aspects are challenging to convey in news articles, which predominantly focus on:

- **Corporate Governance Structure** (in orange in Table 9): This increasing focus reflects the recognition of how governance practices impact corporate performance, accountability, and ethical conduct. Media coverage has spotlighted the need for robust governance structures that ensure companies are managed in the interests of all stakeholders, not just shareholders. In response to these concerns, there is a growing emphasis on enhancing board diversity, ensuring that boards of directors include members with varied backgrounds, experiences, and perspectives [63].
- **Employee Relations and Ethics** (blue): This growing focus reflects an understanding of the profound impact that ethical employment practices and positive workplace relationships have on organizational success and sustainability [64], [65]. In recent years, media coverage has brought to light various issues related to workplace ethics and employee treatment, ranging from unfair labor practices to discrimination and harassment [66], [67]. A key aspect of this discussion involves the cultivation of inclusive and respectful workplace cultures that value diversity and equity. The rise of remote and flexible working arrangements, especially highlighted by the COVID-19 pandemic, has further broadened the scope of employee relations, emphasizing the importance of mental health support, work-life balance, and clear communication in maintaining a cohesive and motivated workforce [68].
- **Executive** and **Employee Compensation** (green): This surge in attention underscores the broader implications of compensation structures on corporate ethics, performance, and stakeholder trust [69]. In recent times, media investigations and shareholder advocacy have spotlighted disparities in compensation between top executives and average employees, raising questions about fairness, equity, and corporate values. Such scrutiny often revolves around the ratios of CEO pay to that of median employees, bonus structures, and the alignment of compensation with long-term corporate performance and sustainability goals [70].

VI. CONCLUSION AND FUTURE WORKS

In this paper, we present a comprehensive analysis of ESG concepts and their evolution from 1980 to 2022, with a focus on news documents from the United States and the United Kingdom. Leveraging the Dow Jones Article dataset, our analysis encompasses news from prominent daily newspapers such as The Guardian, The New York Times, and The Times. To conduct the analysis, we initially applied an extraction pipeline to news articles, involving the structuring of extracted data into a KG. The employed approach utilizes advanced information extraction methodologies to distill relevant information from articles into structured statements represented as triples. These triples were aggregated, verified, and used to construct an extensive knowledge graph. The adopted pipeline is versatile and applicable to any domain, facilitating the analysis of various entity types and establishing semantic relationships between them based on the information extracted from news articles. The information

[20]Ending child labour, forced labour and human trafficking in global supply chains - https://www.ilo.org/wcmsp5/groups/public/—ed_norm/—ipec/documents/publication/wcms_728062.pdf.

extraction pipeline underwent a rigorous evaluation by three annotators, resulting in an accuracy of 0.85.

The resulting knowledge graph was utilized to examine the three fundamental elements of Environmental, Social, and Governance (ESG), uncovering the key topics prevalent in recent news coverage. In the environmental sector, the investigation underscored a strong focus on climate change, the adoption of renewable energy, and the preservation of biodiversity. Regarding the social dimension, the analysis brought to light the growing importance of issues such as racism, gender identity, human rights, the increasing role of charities, and the ethical challenges of modern supply chains. Finally, in the governance area, the research highlighted a focus on issues related to corporate governance accountability, ethics in the workplace, and the behavior and compensation of executives.

In future research, we aim to incorporate domain-specific ontologies or knowledge bases that could aid in establishing more accurate and contextually relevant relationships. We also aim to study the ability of large language models [71] to analyse media and extract structured information from the news. Finally, the exploration of temporal aspects in relation detection, capturing the dynamic evolution of connections over time, would also contribute to a richer understanding of the changing landscape of ESG concepts.

REFERENCES

[1] LISI. (Oct. 2023). *U.K. and EU—An Analysis of ESG Reporting Requirements and Trends*. [Online]. Available: https://www.lisi-law.eu/resources/uk-and-eu-an-analysis-of-esg-reporting-requirements-and-trends

[2] R. Barkemeyer, P. Givry, and F. Figge, "Trends and patterns in sustainability-related media coverage: A classification of issue-level attention," *Environ. Planning C, Politics Space*, vol. 36, no. 5, pp. 937–962, 2018.

[3] A. L. Opdahl, T. Al-Moslmi, D.-T. Dang-Nguyen, M. G. Ocaña, B. Tessem, and C. Veres, "Semantic knowledge graphs for the news: A review," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–38, Dec. 2022, doi: 10.1145/3543508.

[4] F. Anting Tan, D. Paul, S. Yamaura, M. Koji, and S.-K. Ng, "Constructing and interpreting causal knowledge graphs from news," 2023, *arXiv:2305.09359*.

[5] L. Ehrlinger and W. Wöß, "Towards a definition of knowledge graphs," in *Proc. Joint Posters Demos Track, 12th Int. Conf. Semantic Syst. (SEMANTiCS), 1st Int. Workshop Semantic Change Evolving Semantics*, in CEUR Workshop Proceedings, vol. 1695, Leipzig, Germany, M. Martin, M. Cuquet, and E. Folmer, Eds., Sep. 2016, pp. 1–4. [Online]. Available: https://ceur-ws.org/Vol-1695/paper4.pdf

[6] C. Peng, F. Xia, M. Naseriparsa, and F. Osborne, "Knowledge graphs: Opportunities and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13071–13102, 2023.

[7] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, 2015.

[8] D. Dessì, F. Osborne, D. R. Recupero, D. Buscaldi, and E. Motta, "Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain," *Future Gener. Comput. Syst.*, vol. 116, pp. 253–264, Mar. 2021.

[9] D. Dessí, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, and E. Motta, "CS-KG: A large-scale knowledge graph of research entities and claims in computer science," in *Proc. Int. Semantic Web Conf.* Cham, Switzerland: Springer, 2022, pp. 678–696.

[10] F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, M. Simon, S. Villata, and M. Winckler, "Covid-on-the-web: Knowledge graph and services to advance COVID-19 research," in *Proc. 19th Int. Semantic Web Conf.*, Athens, Greece. Cham, Switzerland: Springer, Nov. 2020, pp. 294–310.

[11] A. Chessa, G. Fenu, E. Motta, F. Osborne, D. Reforgiato Recupero, A. Salatino, and L. Secchi, "Data-driven methodology for knowledge graph generation within the tourism domain," *IEEE Access*, vol. 11, pp. 67567–67599, 2023.

[12] M. Rizun, "Knowledge graph application in education: A literature review," *Acta Universitatis Lodziensis. Folia Oeconomica*, vol. 3, no. 342, pp. 7–19, 2019.

[13] A. Tchechmedjiev, P. Fafalios, K. Boland, M. Gasquet, M. Zloch, B. Zapilko, S. Dietze, and K. Todorov, "ClaimsKG: A knowledge graph of fact-checked claims," in *Proc. 18th Int. Semantic Web Conf.*, Auckland, New Zealand. Cham, Switzerland: Springer, Oct. 2019, pp. 309–324.

[14] F. Barile, F. Ricci, M. Tkalcic, B. Magnini, R. Zanoli, A. Lavelli, and M. Speranza, "A news recommender system for media monitoring," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 132–140.

[15] P. Charles, D. de Antonio Liedo, M. Maggi, and J. Palate, "Macroeconomic monitoring and visualizing news," *SSRN J.*, Jan. 2014.

[16] A. Odone and C. Signorelli, "What are we told? A news media monitoring model for public health and the case of vaccines," *Eur. J. Public Health*, vol. 26, no. 4, pp. 533–534, Aug. 2016, doi: 10.1093/eurpub/ckw002.

[17] K. Ma, Z. Yu, K. Ji, and B. Yang, "Stream-based live public opinion monitoring approach with adaptive probabilistic topic model," *Soft Comput.*, vol. 23, no. 16, pp. 7451–7470, 2019.

[18] N. Panagiotou, A. Saravanou, and D. Gunopulos, "News monitor: A framework for exploring news in real-time," *Data*, vol. 7, no. 1, p. 3, Dec. 2021. [Online]. Available: https://www.mdpi.com/2306-5729/7/1/3

[19] J. Ramos, "Using TF-IDF to determine word relevance in document queries," in *Proc. 1st Instructional Conf. Mach. Learn.*, 2003, vol. 242, no. 1, pp. 29–48.

[20] I. Dagan, R. Feldman, and H. Hirsh, "Keyword-based browsing and analysis of large document sets," in *Proc. Symp. Document Anal. Inf. Retr. (SDAIR)*, Las Vegas, NV, USA, 1996.

[21] I. Skibina, M. Dilai, and S. Druzhbiak, "Keyword analysis as a means of identifying dominant topics in news discourse," in *Proc. IEEE 17th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Nov. 2022, pp. 115–118.

[22] S. D. Reese. (Aug. 2016). *Theories of Journalism*. [Online]. Available: https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-83

[23] L. Lloyd, D. Kechagias, and S. Skiena, "Lydia: A system for large-scale news analysis," in *Proc. Int. Symp. String Process. Inf. Retr.* Cham, Switzerland: Springer, 2005, pp. 161–166.

[24] L. Lloyd, P. Kaulgud, and S. Skiena, "Newspapers vs. blogs: Who gets the scoop?" in *Proc. AAAI Spring Symp., Comput. Approaches Analyzing Weblogs*, 2006, pp. 117–124.

[25] L. Lloyd, A. Mehler, and S. Skiena, "Identifying co-referential names across large corpora," in *Combinatorial Pattern Matching*, M. Lewenstein and G. Valiente, Eds. Berlin, Germany: Springer, 2006, pp. 12–23.

[26] D. Vrandečić and M. Krötzsch, "Wikidata: A free collaborative knowledgebase," *Commun. ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[27] J. Piskorski, V. Zavarella, M. Atkinson, and M. Verile, "Timelines: Entity-centric event extraction from online news," in *Proc. Text2Story@ ECIR*, 2020, pp. 105–114.

[28] R. Navigli and S. P. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network," *Artif. Intell.*, vol. 193, pp. 217–250, Dec. 2012.

[29] A. Scharl and A. Weichselbraun, "An automated approach to investigating the online media coverage of U.S. presidential elections," *J. Inf. Technol. Politics*, vol. 5, no. 1, pp. 121–132, 2008.

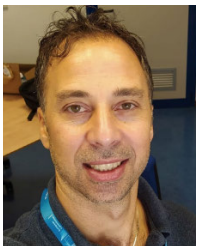
[30] H. Tanev, J. Piskorski, and M. Atkinson, "Real-time news event extraction for global crisis monitoring," in *Natural Language and Information Systems*, E. Kapetanios, V. Sugumaran, and M. Spiliopoulou, Eds. Berlin, Germany: Springer, 2008, pp. 207–218.

[31] A. T. Valero, M. M. Y. Gómez, and L. V. Pineda, "Using machine learning for extracting information from natural disaster news reports," *Computación Sistemas*, vol. 13, no. 1, pp. 33–44, 2009.

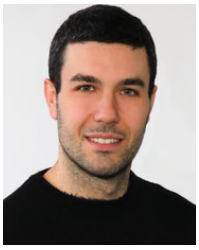
[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[33] S. R. Bowman, "Eight things to know about large language models," 2023, *arXiv:2304.00612*.
[34] P. Hitzler, "A review of the semantic Web field," *Commun. ACM*, vol. 64, no. 2, pp. 76–83, 2021.
[35] S. Gangopadhyay, K. Boland, D. Dessì, S. Dietze, P. Fafalios, A. Tchechmedjiev, K. Todorov, and H. Jabeen, "Truth or dare: Investigating claims truthfulness with claimskg," in *Proc. 2nd Int. Workshop Linked Data-Driven Resilience Res., Extended Semantic Web Conf. (ESWC)*, in CEUR Workshop Proceedings, vol. 3401, 2023.
[36] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. McKeown, and T. B. Hashimoto, "Benchmarking large language models for news summarization," 2023, *arXiv:2301.13848*.
[37] J. Sun, Y. Liu, J. Cui, and H. He, "Deep learning-based methods for natural hazard named entity recognition," *Sci. Rep.*, vol. 12, no. 1, p. 4598, 2022.
[38] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 2, p. e3767, 2020.
[39] F. Al-Obeidat, O. Adedugbe, A. B. Hani, E. Benkhelifa, and M. Majdalawieh, "Cone-KG: A semantic knowledge graph with news content and social context for studying COVID-19 news articles on social media," in *Proc. 7th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*, Dec. 2020, pp. 1–7.
[40] D. Liu, T. Bai, J. Lian, G. Sun, W. X. Zhao, J. R. Wen, and X. Xie, "News graph: An enhanced knowledge graph for news recommendation," in *Proc. KaRS@CIKM*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:204777685
[41] M. Rospocher, M. Van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, and T. Bogaard, "Building event-centric knowledge graphs from news," *J. Web Semantics*, vol. 37, pp. 132–151, Mar. 2016.
[42] L. Fu, H. Peng, and S. Liu, "KG-MFEND: An efficient knowledge graph-based model for multi-domain fake news detection," *J. Supercomput.*, vol. 76, no. 16, pp. 18417–18444, 2023.
[43] D. Dessí, F. Osborne, D. R. Recupero, D. Buscaldi, and E. Motta, "SCICERO: A deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain," *Knowl.-Based Syst.*, vol. 258, Dec. 2022, Art. no. 109945.
[44] A. A. Salatino, F. Osborne, and E. Motta, "How are topics born? Understanding the research dynamics preceding the emergence of new areas," *PeerJ Comput. Sci.*, vol. 3, p. e119, Jun. 2017.
[45] M. Kabir, U. E. Habiba, W. Khan, A. Shah, S. Rahim, P. R. D. los Rios-Escalante, Z.-U.-R. Farooqi, L. Ali, and M. Shafiq, "Climate change due to increasing concentration of carbon dioxide and its impacts on environment in 21st century; A mini review," *J. King Saud Univ. Sci.*, vol. 35, no. 5, 2023, Art. no. 102693. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1018364723001556
[46] A. Slameršak, G. Kallis, and D. W. O'Neill, "Energy requirements and carbon emissions for a low-carbon energy transition," *Nature Commun.*, vol. 13, no. 1, p. 6932, Nov. 2022, doi: 10.1038/s41467-022-33976-5.
[47] A. Opoku, "Biodiversity and the built environment: Implications for the sustainable development goals (SDGs)," *Resour., Conservation Recycling*, vol. 141, pp. 1–7, Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921344918303768
[48] D. Devakumar, S. Selvarajah, I. Abubakar, S.-S. Kim, M. McKee, N. S. Sabharwal, A. Saini, G. Shannon, A. I. R. White, and E. T. Achiume, "Racism, xenophobia, discrimination, and the determination of health," *Lancet*, vol. 400, no. 10368, pp. 2097–2108, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673622019729
[49] A. Edmans, C. Flammer, and S. Glossner, "Diversity, equity, and inclusion," Nat. Bur. Econ. Res., Tech. Rep., 2023.
[50] S. Monro, "Non-binary and genderqueer: An overview of the field," *Int. J. Transgenderism*, vol. 20, nos. 2–3, pp. 126–131, 2019.
[51] C. Kleps, "Race, gender, and place: How judicial identity and local context shape anti-discrimination decisions," *Law Soc. Rev.*, vol. 56, no. 2, pp. 188–212, 2022.
[52] M. Williamson, "A global analysis of transgender rights: Introducing the trans rights indicator project (TRIP)," *Perspectives Politics*, pp. 1–20, 2023.
[53] C. Parsell, A. Clarke, and F. Perales, *Charity and Poverty in Advanced Welfare States*. Evanston, IL, USA: Routledge, 2021.
[54] R. Kober and P. J. Thambar, "Coping with COVID-19: The role of accounting in shaping charities' financial resilience," *Accounting, Auditing Accountability J.*, vol. 34, no. 6, pp. 1416–1429, 2021.
[55] J. Bibby, G. Everest, and I. Abbs, "Will COVID-19 be a watershed moment for health inequalities," Health Found., Tech. Rep., Jul. 2020.
[56] A. Rangone and L. Busolli, "Managing charity 4.0 with blockchain: A case study at the time of COVID-19," *Int. Rev. Public Nonprofit Marketing*, vol. 18, no. 4, pp. 491–521, 2021.
[57] A. Bhati and D. McDonnell, "Success in an online giving day: The role of social media in fundraising," *Nonprofit Voluntary Sector Quart.*, vol. 49, no. 1, pp. 74–92, 2020.
[58] S. R. Maier, "News coverage of human rights: Investigating determinants of media attention," *Journalism*, vol. 22, no. 7, pp. 1612–1628, 2021.
[59] O. Martin-Ortega, F. Dehbi, V. Nelson, and R. Pillay, "Towards a business, human rights and the environment framework," *Sustainability*, vol. 14, no. 11, p. 6596, 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/11/6596
[60] P. Goebel, C. Reuter, R. Pibernik, and C. Sichtmann, "The influence of ethical culture on supplier selection in the context of sustainable sourcing," *Int. J. Prod. Econ.*, vol. 140, no. 1, pp. 7–17, 2012.
[61] S. Kim, C. Colicchia, and D. Menachof, "Ethical sourcing: An analysis of the literature and implications for future research," *J. Bus. Ethics*, vol. 152, pp. 1033–1052, Nov. 2018.
[62] C. F. Wright, "Leveraging reputational risk: Sustainable sourcing campaigns for improving labour standards in production networks," *J. Bus. Ethics*, vol. 137, pp. 195–210, Aug. 2016.
[63] R. Goyal, N. Kakabadse, and A. Kakabadse, "Improving corporate governance with functional diversity on FTSE 350 boards: Directors' perspective," *J. Capital Markets Stud.*, vol. 3, no. 2, pp. 113–136, Nov. 2019, doi: 10.1108/jcms-09-2019-0044.
[64] M. Guerci, G. Radaelli, E. Siletti, S. Cirella, and A. B. R. Shani, "The impact of human resource management practices and corporate sustainability on organizational ethical climates: An employee perspective," *J. Bus. Ethics*, vol. 126, pp. 325–342, Jan. 2015.
[65] S. Villegas, R. A. Lloyd, A. Tritt, and E. F. Vengrouskie, "Human resources as ethical gatekeepers: Hiring ethics and employee selection," *J. Leadership, Accountability Ethics*, vol. 16, no. 2, 2019.
[66] Y. Danilwan and I. P. Dirhamsyah, "The impact of the human resource practices on the organizational performance: Does ethical climate matter?" *J. Positive School Psychol.*, vol. 6, no. 3, pp. 1–16, 2022.
[67] D. C. England, *The Essential Guide to Handling Workplace Harassment & Discrimination*. Berkley, CA, USA: Nolo, 2021.
[68] E. Kurtulus, H. Yildirim, S. Birer, and H. Batmaz, "The effect of social support on work-life balance: The role of psychological well-being," *Int. J. Contemp. Educ. Res.*, vol. 10, pp. 239–249, Mar. 2023.
[69] A. A. Sarhan and B. Al-Najjar, "The influence of corporate governance and shareholding structure on corporate social responsibility: The key role of executive compensation," *Int. J. Finance Econ.*, vol. 28, no. 4, pp. 4532–4556, 2023.
[70] O. Lenihan and N. M. Brennan, "Do boards effectively link firm objectives to CEO bonus performance measures?" *J. Manage. Governance*, Oct. 2023, doi: 10.1007/s10997-023-09690-9.
[71] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy, "Challenges and applications of large language models," 2023, *arXiv:2307.10169*.

SIMONE ANGIONI is currently pursuing the Ph.D. degree with the Department of Mathematics and Computer Science, University of Cagliari, supervised by Diego Reforgiato Recupero. He is also a Main Developer of the academia/industry dynamics (AIDA) knowledge graph, an innovative resource for studying the relationship between academia and industry. His research interests include the science of science, scientometrics, information extraction, the semantic web, and robotics.



SERGIO CONSOLI is currently the Scientific Project Leader of European Commission, Joint Research Centre (DG JRC), Ispra, Italy, the Competence Centre on Composite Indicators and Scoreboards, and formerly with the Centre for Advanced Studies on the project: Big data and forecasting of economic developments. Previously, he was a Senior Scientist with the Data Science Department, Philips Research, Eindhoven, The Netherlands, focusing on advancing automated analytical methods used to extract new knowledge from data for HealthTech applications. Other former experiences include Italian Presidency of the Council of Ministers and the National Research Council of Italy. He also provided ICT consultancy services to Isab, the largest oil refinery in the Mediterranean area. His education and scientific experience fall in the areas of data science, operational research, artificial intelligence, knowledge engineering, semantic reasoning, and machine learning. He is the author of several research publications in peer-reviewed international journals, granted EPO and WIPO patents, edited books, and led conferences in the fields of his work. He has also co-edited two books. He is an Associate Editor of IEEE Access and a member of the Editorial Board of *PLOS One* and the *Journal of Big Data*.



DANILO DESSÌ received the master's and Ph.D. degrees from the University of Cagliari. He has been a Researcher at GESIS Leibniz Institute for the Social Sciences, Germany, since October 2022. Previously, he was a Post-Doctoral Researcher/Senior Researcher with the FIZ Karlsruhe—Leibniz Institute for Information Infrastructure and Karlsruhe Institute of Technology (KIT) with Dr. Harald Sack. His Ph.D. thesis was supervised by Prof. Diego Reforgiato Recupero. He has been visiting researchers in the following centers all around the world: Philips Research (Eindhoven, 2016), Center for Data Science NYU (New York City, 2017), Knowledge Media Institute, The Open University (Milton Keynes, 2018), and Laboratoire d'informatique de Paris Nord, University of Paris 13 (Paris, 2019). He is the coauthor of AI-KG and a Developer of the pipeline for its generation. His current research interests include artificial intelligence, knowledge graphs, science of science, deep learning, and natural language processing.

FRANCESCO OSBORNE is currently a Research Fellow with the Knowledge Media Institute, The Open University, U.K., where he leads the Scholarly Data Mining Team. He is also an Assistant Professor with the University of Milano-Bicocca. His research interests include artificial intelligence, information extraction, knowledge graphs, science of science, and semantic web. He has authored more than a 100 peer-reviewed publications in top journals and conferences in these fields. He collaborates with major publishers, universities, and companies in the space of innovation for producing a variety of innovative services for supporting researchers, editors, and research policy makers. He has released many well-adopted resources, such as the computer science ontology and the artificial intelligence knowledge graph.

DIEGO REFORGIATO RECUPERO received the Ph.D. degree in computer science from the University of Naples Federico II, Italy, in 2004. He has been a Full Professor with the Department of Mathematics and Computer Science, University of Cagliari, Italy, since February 2022. From 2005 to 2008, he was a Postdoctoral Researcher with the University of Maryland College Park, USA. He won different awards in his career (such as the Marie Curie International Reintegration Grant, Marie Curie Innovative Training Network, Best Researcher Award from the University of Catania, Computer World Horizon Award, Telecom Working Capital, Startup Weekend, and Best Paper Award). He co-founded six companies within the ICT sector and is actively involved in European projects and research (with one of his companies, he won more than 40 FP7 and H2020 projects). His current research interests include sentiment analysis, semantic web, natural language processing, human–robot interaction, financial technology, and smart grid. He is the author of more than 200 conferences and journal articles in these research fields, with more than 2800 citations.

ANGELO SALATINO received the Ph.D. degree in early detection of research trends. He is currently a Research Associate with the Intelligence Systems and Data Science (ISDS) Group, Knowledge Media Institute (KMi), The Open University. In particular, his project aimed at identifying the emergence of new research topics at their embryonic stage. His research interests include the semantic web, network science, and knowledge discovery technologies, with a focus on the structures and evolution of science.

Open Access funding provided by 'Università degli Studi di Cagliari' within the CRUI CARE Agreement