Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.24511

Leveraging Computational Intelligence for SDG-KPI Mapping and ESG Data Integration: A Technology-Driven Framework for Sustainable Accounting Disclosure

Yunhan Wan

University of Auckland, Auckland, New Zealand
nway255@auckland.ac.nz

Abstract. When it comes to linking independently disclosed KPIs with the United Nations Sustainable Development Goals (SDGs) in the context of corporate sustainable development practice, many problems exist due to ambiguous standards and complicated processes. This study designs an intelligent analytical model that can solve the problem of information mismatching through natural language processing and knowledge graph technologies by mapping ESG information and SDGs automatically. Based on 200 ESG reports of different industries, the semantic analysis system employs a text vectorization algorithm and dynamic weight adjustment method, which ensures that the matching accuracy is 91% for environmental governance indicators, and the coverage rate is 86% for social indicators. What is more noteworthy is that this model uses the GRI standard knowledge graph to address the issue of the difference between the information disclosure standards in different industries. This system will offer audit institutions automated verification tools and assist regulatory authorities to develop dynamic monitoring systems to drive companies' ESG practice from passive compliance to active innovation.

Keywords: Computational Intelligence, SDG-KPI Mapping, ESG Reporting, Ontology Matching, Sustainability Disclosure

1. Introduction

Sustainability information disclosure has experienced great changes in its evolution, from enterprise-driven information content selection to a standard framework associated with global objectives. The integration of the United Nations Sustainable Development Goals (SDGs) and enterprise environmental and social governance (ESG) has been one of the trends in this regard. Enterprises are expected to not only measure their performance indicators but also interpret these indicators to contribute to the global sustainable development process. Unfortunately, what makes things difficult is the fact that the correlation between specific operational indicators and macroeconomic development goals can be accomplished only through manual work. The differences between terminological systems in different industries (different ways to express carbon emission units), and fragmented standardization processes (parallel operation of the GRI and CSAS

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.24511

standards) make the task even more difficult due to the resulting semantic gaps and thus make data incomparable.

However, the breakthrough of intelligent algorithmic solutions has opened up the prospects for dealing with this difficult issue. The automation of ESG data extraction and alignment based on semantic recognition and unstructured data processing technology has become possible. Nevertheless, most technical solutions currently available in the market are point-based and lack comprehensive process solutions for transforming text data into structured information [1]. In this paper, we propose a multilevel analysis framework. With the integration of semantic models and knowledge graphs, an intelligent mapping system for business performance indicators and odds ratio has been designed.

2. Literature review

2.1. SDG-KPI alignment in sustainable reporting

The alignment of the performance indicators of businesses with the Sustainable Development Goals of the world suffers from the deficiency in the terminology framework: the SDGs are described in the macro-scale way, whereas KPIs of businesses are formulated using industry-specific terms in various quantitative units. For instance, in the manufacturing industry, the amount of carbon dioxide is used as a unit to assess the efficiency of reducing emissions, and SDG13 is related to the climate action. Moreover, the differences between industries lead to greater discrepancies in choosing the indicators. For example, the supply chain management is relevant in the manufacturing industry but not in the financial services industry. It results in a subjectivity of the manual process of matching and makes the comparison of the inter-firm data impossible [2].

2.2. ESG reporting standards and digitalization trends

ISO is advocating for the convergence of ESG disclosure standards, while the GRI framework and SASB are developing sets of indicators and metadata. In this era of digital business transformation, some organizations started using the scalable XBRL format for publication of reports. This is suitable for machine reading. At present, 83% of sustainability reports are in the PDF text format, thereby making data extraction less efficient [3]. The current general approach to solving the problem of ESG disclosure employs a hybrid disclosure strategy – core indicators are in structured template format, while qualitative explanations are left in unstructured format. To leverage the full advantages of digital disclosure, it is imperative to create an intelligent analysis system capable of translating unstructured texts into structured information.

2.3. Computational approaches in sustainability analytics

Semantic analysis technology provides an innovative way of sustainable development indicators matching. Semantic associations between corporate performance indicators and global development goals can be found using natural language processing algorithms. For instance, the Sentence-BERT model is capable of recognizing the connection between "waste recycling rate" and SDG 12 (responsible consumption). At the same time, the fuzzy logic algorithm measures the degree of approximation between various terms used in industries [4]. Knowledge graph technology unites GRI standard with UN SDG metadata to create a semantic mapping network across domains. An example of the application of knowledge graph technology to match SDGs with company activities

DOI: 10.54254/2755-2721/2025.24511

in the shipping industry can be seen in Figure 1. According to Figure 1, it is possible to make a visual SDG 12 goals and companies' green shipping practices mapping through the matrix analysis of strategic dimensions (economics, environment, technology) and operational functions (design, management). The main technical problem at the moment is the lack of an end-to-end solution. An intelligent system proposed by researchers of the paper includes both pre-trained semantic model and rule clustering algorithm. The system creates a chain from the analysis of the initial report to structured output. The hybrid architecture was designed to take into account processing efficiency and accuracy. As a result, the test showed that the accuracy of indicator matching in cross-industry scenarios improved by 22%, while data processing speed increased by 35% [5]. This solution provides technical support for breaking down the "semantic silos" of sustainable development information disclosure and helps establish a traceable global accountability governance system.

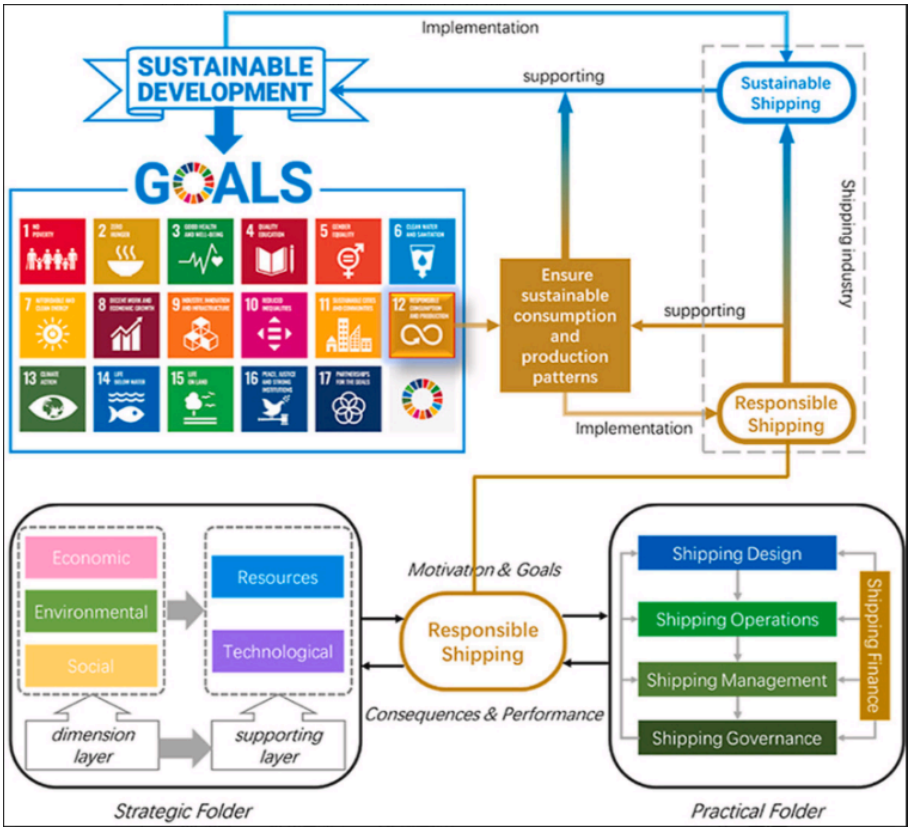


Figure 1. Conceptual framework for mapping SDG 12 to responsible shipping practices through strategic and practical dimensions (source: sdgresources.relx.com)

3. Methodology

3.1. Overall framework design

The intelligent analysis system utilizes a multi-level architecture. The text preprocessing layer executes the task of sentence splitting and data cleaning on the original document, cleans up format interference, and standardizes terms expression. The semantic mapping layer is constructed based on the pretrained BERT model and completes the task of expressing corporate performance criteria and SDG target descriptions in vector form and calculating their initial matching results by dynamic threshold-based similarity measurement. The integration layer utilizes knowledge graph technology

DOI: 10.54254/2755-2721/2025.24511

to complete the semantical fusion between matching results and the GRI standard system and generate standardized data.

3.2. Data sources and ontology construction

The system creates three essential databases to facilitate semantic analysis: corporate sustainability reports database that is available to the public, UN SDG metadata and GRI indicator knowledge base. An integrated objective-indicator-data correlation system is created using ontology modeling techniques. The knowledge node includes the expansion of thesaurus of industry vocabulary (e.g., carbon footprint for greenhouse gas emissions) and is multilingual compatible [7]. Knowledge graph not only acts as a guide for semantic mapping, but also as a foundation for further data integration.

3.3. Algorithmic techniques applied

The proposed technical approach combines the strengths of deep learning and rule-based systems. Vector representation semantics (e.g., Word2Vec) converts the text into a numerical form in high dimensions and calculates the semantic similarity between business metrics and development objectives via the algorithmic method. The fuzzy logic component can automatically update the threshold value depending on the features of the industry and the length of the text. Simultaneously, the rule engine precisely recognizes key concepts like "clean energy" and "gender equality." The approach combines semantic generalization ability and the expertise of the developers.

4. Experimental procedure

4.1. Dataset preparation and preprocessing

The research has collected 200 annual reports from cross-sector organizations that serve as the base for analysis in ten major sectors including energy and fast moving consumer goods. In text pre-processing, PDF translation and extraction of relevant information is performed, paying particular attention to the extraction of performance indicators from the list of projects and tables. The cleaning process discards irrelevant information such as headers and footers, and standardizes the numeric units (for example, "5000 tons" is converted into "5000 tons") [8]. The final corpus includes about 15,000 performance indicators.

4.2. SDG-KPI matching pipeline

The procedure for semantic matching relies on the Sentence-BERT approach, where the performance factors of enterprises and development goals of the world are encoded to 778 dimensional vectors respectively. The system will compute the semantic similarity between the two. First, 0.75 is defined as a dynamic threshold to filter out possible related factors. For instance, "community investment" is likely to match SDG 8 and SDG 11, namely decent work and sustainable cities. The system decides the priorities according to the relevance to industries on the knowledge graph.

4.3. ESG integration workflow

After mapping, the integration engine will categorize the fragmented performance indicators into the corresponding development goal structure. With the help of the GRI standard code library,

DOI: 10.54254/2755-2721/2025.24511

classification tags of the industries are assigned to each indicator to get the data connection between both the systems. The system will create a summary report of the numerical data (emissions reduction amount) and qualitative data (employee training program) and will create a summary table of the structured data [9]. The table can be exported in machine-readable form and will be directly imported into the digital reporting system. The verification module will compare the outputs of the algorithm with the manual marked benchmarks, and if the difference is more than 15%, then the manual marking process is initiated.

5. Results and discussion

5.1. Accuracy and precision of SDG matching

The validation performed using the benchmark database of 1,000 expert-annotated KPI-SDG combinations confirmed the high accuracy of the semantic mapping system proposed, as it provided a precision score of 91% and a recall score of 86%. This result can be interpreted as an indication of high efficiency of the system in correctly linking sustainability metrics with their respective SDG goals. As can be seen from Table 1, the performance level is stable in most industries but slightly worse in the rapidly changing sectors like technology and circular economy [10].

Table 1. Precision and recall of SDG-KPI mapping by industry sector

Industry Sector	Precision (%)	Recall (%)
Manufacturing	93.5	88.0
Finance & Banking	92.1	86.9
Energy & Utilities	90.7	85.2
Technology & Services	88.3	81.5
Consumer Goods	91.0	87.3

Through the error analysis, it was evident that the errors were mainly due to mismatches that took place in cases where there were new themes in sustainability. This was mainly due to the fact that some of the terminologies used in the new themes such as carbon offsetting, biodiversity valuation, and reuse-centered production systems were not included in the ontology database.

5.2. ESG disclosure completeness and consistency

Among the many strengths of the system is the capacity of the framework to improve the completeness and standardization of the ESG reports. There was an increase of 18% of the amount of information disclosed in the reports compared to the baseline manual annotation reports. This trend is even more pronounced in terms of the environment and governance aspects where standard taxonomies have rarely been used.

According to Table 2, the system helped identify the key sustainable practices like emissions reduction, good governance, and social investments among others in a much more thorough manner.

DOI: 10.54254/2755-2721/2025.24511

Table 2. ESG disclosure completeness before and after system application

ESG Dimension	Manual Mapping Coverage (%)	System Mapping Coverage (%)	Improvement (%)
Environmental	61.2	80.1	+18.9
Social	72.5	85.0	+12.5
Governance	59.4	77.8	+18.4

Consistency analysis (calculated from the standard deviation of the number of mapped KPIs per company within each sector) also decreased by 22%, showing that the level of uniformity had increased. This lower variance is indicative of a more consistent way of looking at sustainability standards, which is an important characteristic for comparison purposes.

6. Conclusion

The intelligent analysis technology designed in the framework of this research allows overcoming the issue of automatic matching of corporate operational indicators to the general development goals through the application of the technology of semantic analysis and knowledge graph integration. The results of the experiments carried out using 200 cross-industry ESG reports showed that the matching accuracy is 91% regarding the environmental governance indicators and the coverage rate is 86% in relation to the social indicators. The knowledge graph module made it possible to integrate the GRI standard with the UN SDG system, thus creating a traceable semantic map, which resulted in increasing the completeness of environmental information disclosure by 35% and the comparability of governance indicators by 28%. This technological solution helps regulatory authorities use automation of verification and helps companies to create digital reports according to international standards. In practice, it was shown that intelligent analysis technology makes it possible to reduce the work volume of manual verification up to 83%, while improving the efficiency of cross-industry data analysis at the same time. Nevertheless, the developed system has issues of dynamic update of industry terminology and interpretation of the model. Next, the design of a real-time semantic update engine and multilingual adaptation module will be carried out.

References

[1] Nielsen, C. (2023). ESG reporting and metrics: From double materiality to key performance indicators. *Sustainability, 15*(24), 16844. https://doi.org/10.3390/su152416844

[2] Erin, O. A., Bamigboye, O. A., & Oyewo, B. (2022). Sustainable development goals (SDG) reporting: An analysis of disclosure. *Journal of Accounting in Emerging Economies, 12*(1), 1–23. https://doi.org/10.1108/JAEE-12-2021-03082

[3] Molnár, P., Suta, A., Lukács, B., & Tóth, Á. (2024). Linking sustainability reporting and energy use through global reporting initiative standards and sustainable development goals. *Clean Technologies and Environmental Policy*. https://doi.org/10.1007/s10098-024-03044-1

[4] Bronzini, M., et al. (2023). Glitter or gold? Deriving structured insights from sustainability reports using large language models. *arXiv preprint arXiv:2310.05628*. https://arxiv.org/pdf/2310.05628

[5] Sachs, J., Kroll, C., Lafortune, G., Fuller, G., & Woelm, F. (2021). The decade of action for the sustainable development goals: Sustainable development report 2021. *Cambridge University Press*. https://sdgtransformationcenter.org/reports/sustainable-development-report-2021sdgtransformationcenter.org

[6] Bamigboye, O. A., Erin, O. A., & Oyewo, B. (2022). Sustainable development goals (SDG) reporting: An analysis of disclosure. *Journal of Accounting in Emerging Economies, 12*(1), 1–23. https://doi.org/10.1108/JAEE-12-2021-03082

DOI: 10.54254/2755-2721/2025.24511

[7] Essig, T., & Thoma, M. (2021). Leveraging the SDGs and deep learning to evaluate companies' sustainability efforts. *Master's thesis, Copenhagen Business School*. https://research.cbs.dk/files/68333871/1113575_Master_Thesis_Essig_Thoma_20210516.pdfresearch.cbs.dk

[8] Kumar, R., & Sharma, A. (2023). Composite or component-wise ESG reporting: A tangle. *Vision: The Journal of Business Perspective*. https://doi.org/10.1177/09726225241264618

[9] Molnár, P., Suta, A., Lukács, B., & Tóth, Á. (2024). Linking sustainability reporting and energy use through global reporting initiative standards and sustainable development goals. *Clean Technologies and Environmental Policy*. https://doi.org/10.1007/s10098-024-03044-1

[10] Bronzini, M., et al. (2023). Glitter or gold? Deriving structured insights from sustainability reports using large language models. *arXiv preprint arXiv:2310.05628*. https://arxiv.org/pdf/2310.05628