Automatic Detail Extraction from Sustainability Objectives Using Weak Supervision

Mohammad Mahdavi
Gisma University of Applied Sciences
Potsdam, Germany
mohammad.mahdavi@gisma.com

Tom Debus
Ferris Solutions AG
Cham, Switzerland
tom.debus@ferris.ai

Abstract

Consumers are increasingly pressuring companies to disclose sustainability data as sustainable development gains public importance. In response, companies publish their goals and strategies in various online sustainability reports. Domain experts must continuously analyze these reports to combat greenwashing by extracting key details of companies' sustainability objectives for structured databases, which is time-consuming.

In this paper, we propose a novel approach to automatically extract key details from sustainability objectives using weak supervision. Our method tokenizes the objectives and employs a weakly supervised token-labeling algorithm that converts coarse, objective-level annotations into token-level labels. A transformer model is then fine-tuned on these weak supervision signals for sequence labeling. This approach enables automatic extraction of key details from new sustainability objectives without costly token-level annotation. Experiments show that our method outperforms state-of-the-art approaches for this task. We further integrate our approach into GoalSpotter, an existing sustainability objective detection system, to demonstrate its significant post-deployment impact in real-world business applications.

Keywords

Sustainability Reporting, Greenwashing, Information Extraction, Weak Supervision, Sequence Labeling, Transformer Models

1 Introduction

Sustainable development is crucial for ensuring that current and future generations can thrive by balancing economic growth, social equity, and environmental protection. In 2015, the United Nations adopted the 2030 agenda to call for global cooperation in achieving 17 sustainable development goals, such as ending poverty, reducing inequality, and combating climate change [30].

As a result, sustainability strategies and corporate performance have become more important than ever for the public [8, 29]. Various stakeholders (e.g., investors, consumers, and policymakers) push companies to disclose their environmental/social information transparently [5]. Analyzing the published sustainability reports allows sustainability experts to evaluate companies' performance and spot greenwashing, which is the act of misleading the public with respect to the environmental practices of a company or environmental benefits of a product/service [8]. To this end, domain experts need to extract and fact-check the sustainability claims of companies over time to make sure they are delivering their previously promised goals [14, 15].

Traditionally, domain experts read and analyze sustainability reports of companies manually. In particular, they extract key information from numerous lengthy sustainability reports and store it in structured databases. This process enables them to compare companies in terms of their sustainability goals, track their progress toward these goals, and evaluate their overall sustainability strategies and performance. Naturally, this manual inspection process is not scalable in the big data era [14, 31].

Automatic processing of sustainability reports is therefore an emerging need. Recent research shows that some aspects of this challenging problem, such as detecting sustainability objectives, can be effectively automated. We have already developed GoalSpotter [14] that is a sustainability objective detection system. It fine-tunes transformer models to filter out text blocks in sustainability reports that contain objectives. However, although GoalSpotter identifies sustainability objectives, it cannot extract their key details in a structured format suitable for databases with predefined fields.

In this paper, we focus on the problem of extracting key information from sustainability objectives. We formulate this problem as an information extraction task, using weak supervision to identify key details from each sustainability objective without requiring token-level annotations. In particular, we make the following contributions.

- We introduce a weak supervision algorithm that converts coarse, objective-level annotations into weak token-level labels, enabling the creation of a labeled dataset for sustainability detail extraction without costly expert token-level annotation.
- We propose an information extraction system that leverages these weak supervision signals to extract fine-grained details from sustainability objectives. This new extraction service is fully integrated into our previously developed GoalSpotter system and is now publicly available online[1].
- We extensively evaluate the effectiveness and efficiency of our system in comparison to state-of-the-art sustainability objective detail extraction approaches, demonstrating the superiority of our system on various datasets.
- We deploy our system and report its post-deployment impact in real-world business applications.

2 Foundations

2.1 Sustainability Reports

Sustainability reporting is the process of disclosing companies' sustainability information. Different stakeholders (e.g., investors, consumers, and policymakers) demand these sustainability reports to assess the performance/strategies of companies in terms of their environmental and social impacts [5]. Examples of sustainability documents are CSR (Corporate Social Responsibility) and ESG (Environmental, Social, and Corporate Governance) reports [5, 17].

Example 1 (Sustainability reports). For example, Figure 1 shows a small snippet from a long sustainability report. □

[1] https://github.com/Ferris-Solutions/goalspotter_public

EDBT '26, Tampere (Finland)

Table 1: A few sustainability objectives and their annotated details.

Sustainability Objective	Action	Amount	Qualifier	Baseline	Deadline
We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040.	reach	net-zero	carbon		2040
Restore 100% of our global water use by 2025.	Restore	100%	global water use		2025
Reduce energy consumption by 20% by 2025 (baseline 2017).	Reduce	20%	energy consumption	2017	2025

Climate change is one of the world's greatest crises, and to address it, the public and private sectors need to act together.	**We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040.**
Reducing carbon emissions in transportation is a complex challenge for many companies. Businesses also face the challenge of removing carbon emissions from new building construction.	
Detected Sustainability Objectives: "We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040."	

Figure 1: A small snippet from a sustainability report. GoalSpotter detects the text block that contains an environmental claim.

2.2 Sustainability Objectives

A key part of any sustainability report is the sustainability goals of the company, which should be effectively communicated to business stakeholders [19]. Through these goals, companies aim to reach specific sustainability targets by adapting their strategies. Sustainability objectives are specific claims intended to achieve broader sustainability goals. These objectives typically include the following components:

- **Action** is a verb describing the nature of the intended change.
- **Amount** is a relative or absolute value specifying the magnitude and unit of the change.
- **Qualifier** is a short phrase that provides additional context to the amount value.
- **Baseline** is the date when the change process began.
- **Deadline** is the date by which the change is expected to be completed.

Example 2 (**Sustainability objectives**). For example, Table 1 shows a few sustainability objectives and their annotated details. □

2.3 Sustainability Objective Detection

Previous research has shown that the sustainability objective detection problem can be formulated as a text classification task, where text blocks within reports are classified into objective and noise classes [14]. GoalSpotter [14] is one such system that detects sustainability objectives in sustainability reports.

Example 3 (**Sustainability objective detection**). For example, as Figure 1 shows, GoalSpotter detects the bold text block as a sustainability objective. □

2.4 Problem Formulation

To further automate the evaluation of sustainability objectives, we aim to extract their key details from detected objectives and store them in structured databases. Suppose $O = \{o_1, o_2, \ldots, o_{|O|}\}$ is a set of historical sustainability objectives. Let $V_i = \{v_{\text{action}}, v_{\text{amount}}, v_{\text{qualifier}}, v_{\text{baseline}}, v_{\text{deadline}}\}$ be the set of annotated details for the sustainability objective o_i. Given a new sustainability objective o_{new}, the problem is to automatically extract the details of the given sustainability objective, i.e., $V_{\text{new}} = \{v_{\text{action}}, v_{\text{amount}}, v_{\text{qualifier}}, v_{\text{baseline}}, v_{\text{deadline}}\}$.

Automatically extracting structured details from sustainability objectives provides several practical benefits. While the full objective text is always needed for complete interpretation, representing key components, such as *Action*, *Amount*, *Qualifier*, *Baseline*, and *Deadline*, in a structured form enables more efficient indexing, filtering, and partial comparison across large collections of objectives. In particular, fields like *Baseline* and *Deadline* allow tracking progress over time and monitoring whether companies fulfill their stated commitments. Moreover, this structured representation lays the foundation for future extensions, such as normalization or categorization of actions and amounts, which could support more fine-grained analysis and benchmarking across companies.

Example 4 (**Sustainability objective detail extraction**). For example, given the detected sustainability objectives in Table 1, the goal is to extract their key details, namely action, amount, qualifier, baseline, and deadline. □

3 Automatic Detail Extraction

3.1 System Overview

Figure 2 illustrates the workflow of our system for extracting key details from sustainability objectives. Given a collection of detected objectives, the system extracts their relevant information using a two-phase process: a development phase (purple) and a production phase (blue).

The development phase focuses on training a sustainability detail extraction model using weak supervision. In step 1, the system tokenizes the texts of sustainability objectives along with the corresponding domain expert annotations to generate token sequences. In step 2, the token sequences of objectives and annotations are aligned to assign weak token-level labels, identifying tokens that correspond to key details. Finally, in step 3, the weakly labeled token sequences are used to fine-tune a transformer model, enabling it to learn from partial supervision rather than requiring full token-level annotations.

The production phase applies the trained model to extract details from new sustainability objectives. In step 1, the text of a new objective is tokenized into sequences. In step 2, the fine-tuned transformer model predicts a label for each token, determining whether it represents a key detail. This workflow demonstrates how weak supervision allows effective training of sequence-labeling models with minimal annotation effort.

3.2 Weakly Supervised Token Labeling

Information extraction models require standard datasets, such as CONLL-2003 [25], which are labeled at the token level. In these datasets, each token is annotated with a label in a standard format, such as IOB that marks the beginning (B), inside (I), and outside (O) of each entity type.

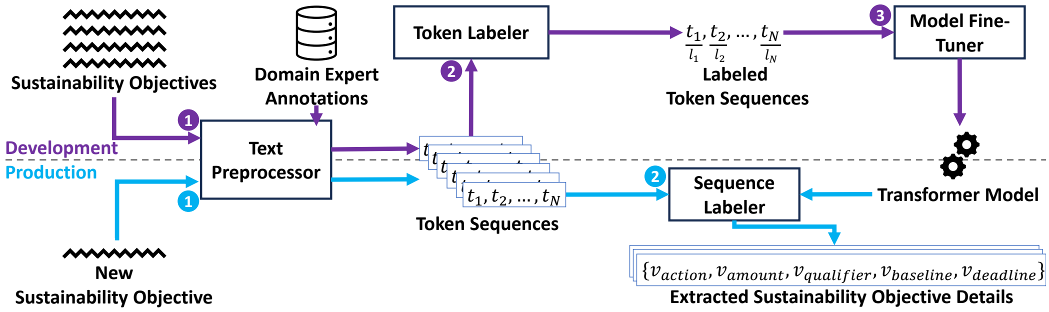


Figure 2: The workflow of our system.

Table 2: An example of a token-level labeled sentence.

Token	Label
Albert	B-PER
Einstein	I-PER
was	O
born	O
in	O
Germany	B-LOC
.	O

We co-founded The Climate Pledge, a commitment to reach net-zero carbon by 2040.

Annotated Key Details: {"Action": **"reach"**, "Amount": **"net-zero"**, "Qualifier": **"carbon"**, "Baseline": **""**, "Deadline": **"2040"**}

Figure 3: An annotated training sustainability objective.

Example 5 (Token-level labels). For example, Table 2 shows a short sentence from the CONLL-2003 dataset [25], where each token is labeled in IOB format. □

Such a token-level labeling process is time-consuming and not scalable in the sustainability domain, where domain experts deal with thousands of heterogeneous sustainability reports every day [14]. That is why the input sustainability objectives to our system are only partially annotated per training instance. That is, each input training sustainability objective might have some annotated key-value pairs that show which key details are mentioned in the given training instance.

Example 6 (Sustainability objective-level annotations). For example, Figure 3 shows a short sustainability objective, which is annotated with some key-value pairs. Note that in real-world data, the sustainability objective might be much longer and the annotations might not contain all key details. □

Therefore, we design a *weak supervision* [22] algorithm that converts coarse sustainability objective-level annotations into token-level labels, enabling sequence labeling models to be trained from partial supervision [3]. As shown in Algorithm 1, given a sustainability objective o, the input text is first tokenized into a sequence of tokens $T = [t_1, \dots, t_{|T|}]$. Initially, all tokens in T are weakly labeled as 'O' (i.e., *outside*), producing $L = [O, \dots, O]$ where $|L| = |T|$. For each annotated key-value pair associated with objective o, the value v is tokenized into a sequence of tokens $U = [u_1, \dots, u_{|U|}]$. The algorithm then searches for the starting position s of the tokenized substring U within T. If a match is found, the corresponding tokens in T are assigned IOB-style weak labels: the first token of U is labeled as B-k (the *beginning* of entity type k), while the remaining tokens are labeled as I-k (the *inside* of entity type k). This procedure produces token-level

Algorithm 1: WeakSupervisionTokenLabeling(o, A)

Input: sustainability objective o, set of annotated details A.
Output: list of weak token labels L.

$T \leftarrow$ tokenize the sustainability objective o into $[t_1, \dots, t_{|T|}]$;
$L \leftarrow$ initialize weak token labels to $[\text{O}, \dots, \text{O}]$, where $|L| = |T|$;
foreach (k, v) *in annotation set* A **do**
  $U \leftarrow$ tokenize the annotation value v into $[u_1, \dots, u_{|U|}]$;
  $s \leftarrow$ find an index s in T where $T[s : s + |U|] = U$;
  if $s \neq -1$ **then**
    $L[s] \leftarrow$ B-k (weak label);
    for $i \leftarrow 1$ *to* $|U| - 1$ **do**
      $L[s + i] \leftarrow$ I-k (weak label);
return L as weak supervision signals;

Table 3: The weakly supervised token-labeling algorithm converts objective-level annotations into token-level labels for each sustainability objective.

Annotated Key Details:
{"Action": **"reach"**, "Amount": **"net-zero"**, "Qualifier": **"carbon"**, "Baseline": **""**, "Deadline": **"2040"**}

Token	Label
We	O
co	O
-	O
founded	O
The	O
Climate	O
Pledge	O
,	O
a	O
commitment	O
to	O
reach	B-Action
net	B-Amount
-	I-Amount
zero	I-Amount
carbon	B-Qualifier
by	O
2040	B-Deadline
.	O

supervision consistent with the coarse objective-level annotations, effectively transforming partial labels into sequence-level training signals. In this sense, our method works as a form of weak supervision and partial label learning [3, 22], where inexpensive coarse annotations are leveraged to train fine-grained sequence models.

Example 7 (Weak supervision token labeling algorithm). For example, Table 3 shows the output of the weak supervision token labeling algorithm for the short annotated sustainability objective in Figure 3. Using the key-value annotations at the objective level, the algorithm assigns a label to each token of the sustainability objective. □

In real-world sustainability reports, objectives are often noisy, incomplete, and heterogeneous, reflecting differences in reporting styles, terminology, and levels of detail across organizations. Our workflow is designed to operate under these realistic conditions by relying on partial, objective-level annotations and weak supervision, rather than assuming complete or clean token-level labels. Following the preprocessing strategy used in GoalSpotter [14], we normalize input texts and remove unnecessary characters to reduce superficial noise. We then apply standard subword tokenization mechanisms provided by modern transformer encoders, such as Byte-Pair Encoding [27], which represent the state-of-the-art in many natural language processing tasks and offer robustness to rare words, morphological variation, and domain-specific terminology.

3.3 Sequence Labeling

Once the weakly supervised token-labeled dataset is constructed for the sustainability detail extraction task, we train a sequence-labeling model. We fine-tune transformer-based encoders for this task, where the model learns to assign token labels using contextual information while being trained on weak supervision signals.

As shown in our experiments, we systematically evaluate the impact of different transformer architectures and hyperparameters on extraction performance. In the default configuration of our prototype, we fine-tune a *RoBERTa* model on the weakly labeled dataset for up to 10 epochs, with a learning rate of 5e-5, a batch size of 16, and the *Adam* optimizer. These hyperparameters were selected based on validation performance and are consistent with common practices for transformer-based sequence labeling. This setup demonstrates that effective sequence models can be trained even when only partial supervision is available, avoiding the need for costly token-level expert annotations.

4 Evaluation

Our experiments aim to answer the following research questions.

- How does our system compare to the existing sustainability objective detail extraction approaches?
- How do our design decisions, including data preprocessing, model selection, and hyperparameter tuning, affect the system performance?

We first introduce our experimental setting and then detail our experiments.

4.1 Setup

Dataset. We evaluate our system on 2 datasets.

- *NetZeroFacts* [32] is a climate goal extraction dataset containing > 14000 passages from climate-related business reports. We have extracted 599 sentences from these emission goal passages, each of which is annotated with at least one label, such as *target value*, *reference year*, and *target year*. We use this dataset to evaluate our system based on recent benchmarks.
- *Sustainability Goals* is our own proprietary dataset containing 1106 sustainability objectives. Each sustainability objective is annotated with 5 key fields, including *Action*, *Amount*, *Qualifier*, *Baseline*, and *Deadline*. To create this dataset, we collected 718 sustainability reports from 422 companies. Domain experts identified a few sustainability objectives in each report and annotated them with the above-mentioned key information. We use this highly heterogeneous and imbalanced dataset to evaluate our system in challenging real-world scenarios.

Baselines. We compare our system to 3 baseline approaches.

- *Conditional random fields* is a traditional statistical model that has been applied to information extraction tasks [20]. We train this model with token-level lexical, orthographic, and contextual features to assess the performance of traditional approaches on modern information extraction tasks, such as sustainability objective detail extraction.
- *Zero-shot prompting* is a technique in which a large language model is asked to extract key information from sustainability reports without task-specific examples [9]. Similarly, we prompt the open-weight Llama 4 109B to extract key details from our sustainability objectives. We implement this baseline to check the performance of zero-shot learning approaches for the sustainability objective detail extraction tasks.
- *Few-shot prompting* technique requires the prompt to include some input-output example instructions [32]. Therefore, in accordance with previous research [32], our prompts here to Llama 4 109B include three input sustainability objectives and the desired output extracted details. This way, we evaluate the performance of few-shot learning approaches for the sustainability objective detail extraction tasks.

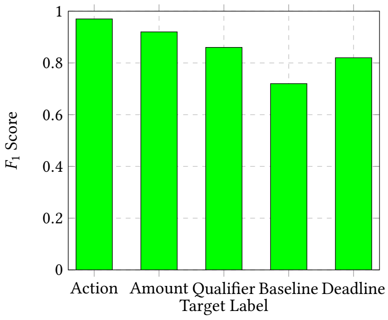
Evaluation measures. We leverage the typical *Precision*, *Recall*, and F_1-*Score* to measure the effectiveness of the information extraction approaches. Formally, $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$, and $F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$, where true positive (TP) is the number of times the approach correctly extracted the information that was actually present, false positive (FP) is the number of times the approach incorrectly extracted information that was or was not actually present, and false negative (FN) is the number of times the approach failed to extract information that was actually present. All evaluation metrics are measured on an unseen test set, which forms 20% of the original dataset. For each evaluation measure, we report the mean of 5 independent runs. For the sake of readability, we omit the standard errors as they are always small numbers close to zero (< 1%).

We also report the training/fine-tuning and inference time in minutes to evaluate the efficiency. We run the experiments on an Ubuntu 18.04 LTS machine with 64 2.5 GHz CPU cores, 256 GB memory, and an NVIDIA RTX A500 4 GB GPU.

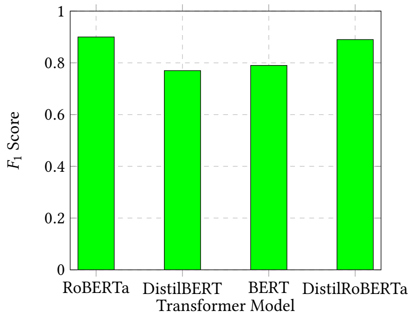
4.2 Comparison with the Baselines

We compare the effectiveness and efficiency of GoalSpotter (which is equipped with our detail extraction approach) with all the baselines in Table 4. GoalSpotter outperforms all the baselines in terms of F_1 score across all the datasets. This demonstrates the value of our weakly supervised token labeling algorithm and careful model fine-tuning, enabling the model to learn effectively from partial annotations across diverse sustainability objectives.

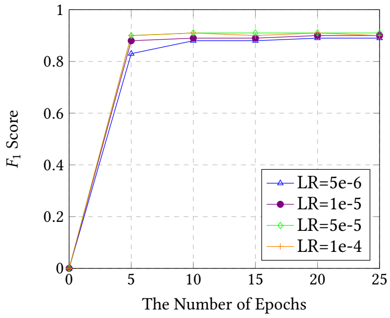
We also observe that our lightweight system, which requires only a few minutes to fine-tune a transformer model, outperforms massive pretrained large language models like Llama 4 109B. This underscores that, with well-designed preprocessing and weak supervision techniques like our token labeling algorithm, complex information extraction tasks can be performed effectively and efficiently without relying on extremely large models.



(a) The Effect of the Target Label



(b) The Effect of the Transformer Model



(c) The Effect of the Epochs Count and Learning Rate

Figure 4: System effectiveness with different internal design decisions on the *Sustainability Goals* dataset.

Table 4: System effectiveness and efficiency in comparison to the baselines. Time is measured in minutes.

Approach	NetZeroFacts				Sustainability Goals			
	P	R	F	T	P	R	F	T
Conditional Random Fields	0.64	0.59	0.61	< 1	0.60	0.86	0.71	< 1
Zero-Shot Prompting	0.63	0.65	0.64	7	0.71	0.86	0.78	13
Few-Shot Prompting	0.70	0.94	0.80	7	0.81	0.96	0.88	13
GoalSpotter	0.87	0.83	**0.85**	2	0.89	0.95	**0.92**	3

4.3 Experiments on Internal Design Decisions

We analyze the effect of the most important design decisions on the effectiveness of our sustainability objective detail extraction system. Figure 4 shows the results of the corresponding experiments with target labels, model selection, and hyperparameters on the *Sustainability Goals* dataset.

The effect of the target label. As shown in Figure 4, the effectiveness of our system on different target labels could vary slightly due to the available amount of labeled data. Our system can achieve a very high F_1 score in extracting *Action* of the sustainability objective because this annotated information is available for 85% of the data points. On the other hand, our system achieves a lower F_1 score when extracting more specific target labels, such as *Baseline* and *Deadline*, since these labels are present in only 14% and 34% of the annotated sustainability objectives, respectively.

The effect of the transformer model. As shown in Figure 4, *RoBERTa* models achieve slightly higher F_1 scores than *BERT* models. Furthermore, the original versions of these models perform slightly better than their distilled versions as expected. This is due to the fact that distilled versions of transformer models are designed to be smaller, faster, and more efficient while achieving almost the same effectiveness as the original model. By default, we use a *RoBERTa* model as it achieves the best effectiveness with an acceptable fine-tuning speed.

The effect of the number of epochs and learning rate. As shown in Figure 4, the number of epochs and learning rate do not affect the model convergence speed significantly if they are chosen from their typical ranges. By setting the learning rate to $5e-5$, our model achieves its highest F_1 score in a few, i.e., 10, epochs.

Table 5: Summary of post-deployment data.

Company	#Documents	#Pages	#Extracted Objectives
C1	20	2131	150
C2	18	3172	642
C3	41	3560	447
C4	19	2488	102
C5	17	1298	113
C6	29	3278	343
C7	23	2208	247
C8	22	5012	764
C9	64	4791	379
C10	16	1202	79
C11	17	1229	95
C12	64	1721	71
C13	18	3250	105
C14	12	2531	43
Total	380	37871	3580

5 Deployment and Demonstration

We demonstrate the post-deployment impact of our detail extraction system for sustainability objectives across two practical scenarios, followed by a brief discussion of the observed results.

5.1 Scenario 1: Detail Extraction from Top Companies' Sustainability Objectives

In our prior research, we have shown how our sustainability objective detection system, GoalSpotter, has been successfully deployed in industry [14]. We deployed GoalSpotter in Ferris Solutions[2], which is a Switzerland-based company focusing on building artificial intelligence solutions for sustainability problems. In short, we ran GoalSpotter on 380 sustainability reports from 14 companies with 37,871 pages and extracted 3,580 sustainability objectives. Table 5 shows the summary of our post-deployment data. We reported the top 2 extracted sustainability objectives per company in our prior research [14].

We have now integrated our new detail extraction service into our previously developed GoalSpotter system to additionally extract details of these already detected sustainability objectives and store them in structured databases. Table 6 shows the extended results from our previous paper, including the details that our system extracts from the same top two sustainability objectives per company.

We observe that our detail extraction approach effectively extracts details of sustainability objectives across various companies and data domains. These results again showcase the effectiveness of our system on new real-world, heterogeneous data.

[2] https://www.ferris.ai

Table 6: The extracted details for the top 2 sustainability objectives per company from the post-deployment data.

Company	Sustainability Objective	Action	Amount	Qualifier	Baseline	Deadline
C1	Integrate sustainability information into their reporting cycle	Integrate		reporting cycle		
C1	Voluntary turnover rate in 2021: 8.1%		8.1%	Voluntary turnover rate		
C2	Substitute F-gases for low GWP alternatives	Substitute		F-gases		
C2	Commitments to double environmental efficiency with new energy, water and waste targets	Commitments	double	environmental efficiency		
C3	We are committed to empowering 100 million smallholder farmers in low to middle	empowering	100 million	smallholder farmers		
C3	Transition all Consumer Health products to 100% recyclable or reusable packaging	Transition	100%	Consumer Health		
C4	Explore innovative value-based approaches	Explore		value-based approaches		
C4	Reduce employees' risk of a serious incident or fatality	Reduce		risk of a serious incident or fatality		
C5	Expand principles of sustainability and performance indicators at key suppliers	Expand		principles of sustainability and performance indicators		
C5	250 students in STEM awareness activities		250	students in STEM awareness activities		
C6	Define sustainability strategies, goals and policies in consultation with key stakeholders	Define		sustainability strategies, goals and policies		
C6	Join industry peers, UN entities and/or other stakeholders in initiatives contributing to solving common challenges and dilemmas at the global and/or local levels	Join		common challenges and dilemmas		
C7	Uses 25 percent PCR content in bottle	Uses	25 percent	PCR content in bottle		
C7	100 percent of major brands share product sustainability information on their websites		100 percent	major brands share product sustainability information		
C8	Promote the proportion of women in leadership positions at the company	Promote		proportion of women in leadership positions at		
C8	Perform waste audits to identify ways to reduce waste or increase recycling efficiency	Perform		waste		
C9	Align strategies, goals and incentive structures of all business units and subsidiaries with corporate sustainability strategy	Align	all	strategies, goals and incentive structures		
C9	Implement water saving programs at the top-10 sites with highest water footprint and water scarcity	Implement		water footprint and water scarcity		
C10	Demonstrate added value of new products	Demonstrate		added value of new products		
C10	Pursue leadership in HAEMOPHILIA	Pursue				
C11	Incorporate environmental sustainability across all aspects of our organization	Incorporate	all	environmental sustainability		
C11	Integrate immunization delivery and family planning services in an effort	Integrate		immunization delivery and family planning services		
C12	30% increase in the representation of women in key leadership roles	increase	30%	representation of women in key leadership roles		
C12	Reached goal of 20% of women in key positions a year ahead of schedule	Reached	20%	women in key positions		
C13	By 2025: priority sites located near sensitive natural areas shall implement a biodiversity protection	implement		biodiversity protection		2025
C13	By 2025, pilot projects will be implemented for promoting further the sustainable use and responsible	will be implemented		sustainable use		2025
C14	Share high-quality medical resources, provide material and technical support to poor areas	Share		medical resource		
C14	Make monthly contributions to the schemes at approximately 7% to 10% of the relevant income	Make	7%	monthly contributions		

Domain experts store these structured data in databases to compare different target companies, monitor their progress toward their sustainability goals, and evaluate companies in terms of their sustainability performance. In particular, they can see in the table that companies, such as C12 and C13, are more specific in terms of indicating the exact amount of change and the timeline of their sustainability objectives than the other companies. At the same time, these specific facts and figures can be monitored over time to measure the fidelity of the companies to their previously claimed sustainability objectives.

5.2 Scenario 2: Detail Extraction from a Single Sustainability Report

We also demonstrate how our detail extraction approach works in a more microscopic scenario, where we analyze one specific sustainability report in detail. We ran GoalSpotter on a single sustainability report to first detect its objectives and then extract their corresponding details, organizing all extracted information into a structured table. Table 7 shows the results for the top sustainability objectives identified in this report.

We observe that our system can effectively extract fine-grained details even when operating on a single document with dense and varied sustainability content. In particular, the extracted details provide a clearer understanding of the company's commitments, the specificity of its planned actions, and the quantitative measures it sets for its sustainability goals. This scenario illustrates that GoalSpotter is not only capable of handling large-scale, multi-company datasets but also performs robustly in focused, report-level analyses. Such document-level insights enable domain experts to perform deeper assessments, verify company claims more precisely, and track the evolution of individual sustainability objectives over time.

5.3 Discussion

Overall, the deployment results demonstrate that our approach reliably extracts meaningful structured details from real-world sustainability objectives across diverse companies, reporting styles, and data domains. The resulting structured information enables large-scale indexing, comparison, and monitoring of sustainability commitments, highlighting the practical utility of our method in post-deployment settings.

Table 7: The extracted details from an example sustainability report.

Sustainability Objective	Action	Amount	Qualifier	Baseline	Deadline
Reduce single-use beverages per seated headcount by 20% relative.	Reduce	20%	single-use beverages per seated headcount		
Keep products and materials in use, and promote healthy materials and safe.	Keep		products and materials		
Achieve Zero Waste to Landfill for our global data center operations.	Achieve	Zero	Waste to Landfill		
10% at our Bay Area headquarters achieving a 15% reduction in landfill waste.	reduction	10%	landfill waste		
Reduce potable water intensity at our Bay Area headquarters by 5% by the end of 2019, against a 2017 baseline.	Reduce	5%	potable water intensity	2017	2019
By 2023, we will install 1 million energy- and money-saving thermostats in homes that need them most.	will install	1 million	energy- and money-saving thermostats		2023

At the same time, the results also reveal several expected limitations when operating on noisy and heterogeneous real-world data. As shown in Tables 6 and 7, some extracted records remain incomplete, which is largely due to the fact that many sustainability objectives omit certain details (e.g., *Baselines* or *Deadlines*) or express them implicitly. In addition, objectives that contain multiple actions or targets within a single sentence may partially confuse the extraction model, leading to missing or fragmented outputs. Finally, our current implementation relies on exact token-level matching between annotations and sustainability objectives, which limits its ability to capture semantically equivalent but lexically different expressions.

Despite these limitations, the extracted structured information remains valuable for large-scale analysis and monitoring, and these observations point to natural directions for future improvements, such as objective segmentation, fuzzy matching, and semantic normalization.

6 Related Work

6.1 Sustainability Objective Detection

Sustainability objective detection aims to identify environmental and social claims of companies, such as "we will reach net-zero carbon by 2040", in their sustainability reports [14]. Existing approaches typically formulate this as either a text classification [14, 28, 31] or text retrieval [2, 7, 21, 32] tasks.

In both task formulations, the sustainability reports are segmented into smaller text units, such as text blocks [7, 14], tweets [31], sentences [2, 28], or passages [32]. In the text classification formulation, each text unit is annotated with labels, such as sustainability objective or noise. Therefore, a classifier, which is usually a transformer model such as RoBERTa [14, 28, 31], learns to identify sustainability objectives. On the other hand, in the text retrieval formulation, the text units are ranked based on their sustainability objective score [2, 7, 32], which represents the keyword/embedding similarity of the text units to some predefined sustainability queries.

Sustainability objective detection is an upstream orthogonal step to our detail extraction task. In fact, any of the above approaches can be first run on the sustainability reports to detect sustainability objectives. Our system can use the outputs of any of these approaches as input to extract detailed information.

6.2 Detail Extraction from Sustainability Objectives

Once the sustainability objectives are identified, the detail extraction task extracts the key information from text units to store them in a structured way. Existing solutions can be categorized into rule-based [6], zero-shot prompting [9], and few-shot prompting [2, 32] approaches.

Rule-based approaches rely on user-provided configuration to extract key information from sustainability reports. easIE [6] is a rule-based semi-automatic framework that extracts sustainability metrics from CSR reports of the companies on the web. It requires the user to provide a configuration file, specifying a CSS selector that points the system to the set of HTML elements that contain the target information [6]. In contrast, our system does not need any given rule as it learns to automatically extract details of sustainability objectives.

In the zero-shot prompting, a large language model, such as GPT-4, analyzes the sustainability reports directly in PDF format to extract key information without being given task-specific instructions [9]. Prior research shows that this approach can be inferior even compared to simple keyword-based search methods, which involve searching for key details using keywords [9]. These results motivate our approach that fine-tunes transformer models for the detail extraction task.

In the few-shot prompting, a large language model, such as GPT-3.5-turbo [32], is instructed with a few input-output examples to extract key details from given sustainability objectives. This key information could contain the objective category (e.g., "waste"), predicate (e.g., "investment in"), object (e.g., "cutting-edge recycling technologies") [2], change percentage (e.g., 10%), target year (e.g., 2030), and reference year (e.g., 2020) [32]. This extracted information is then stored in a knowledge base [2, 32]. The learning capacity of large language models in the few-shot learning scenarios is limited as only a few input-output examples can be used to instruct the model. As a result, existing few-shot learning approaches can only extract details of a narrow category of sustainability objectives, such as only carbon emission [32], only climate [13], or only energy consumption [11] objectives.

In contrast, our fine-tuned model systematically learns to extract details of any category of sustainability objectives in various domains without requiring any predefined rules. That is why our system significantly outperforms zero/few-shot learning approaches, as shown in the experiments.

6.3 Information Extraction

Information extraction is the task of extracting structured key information from unstructured text [4], such as keyphrases [16], named entities [1], relationships [18], and events [33].

Existing techniques can be divided into three categories: (1) rule-based, (2) machine learning-based, and (3) deep learning-based approaches [34]. Rule-based methods require domain experts to design and maintain regular expressions, templates, or dictionaries [12]. Machine learning-based approaches rely on token labels (e.g., *B-PER*, *I-PER*, *O*) to train traditional sequence models, such as hidden Markov models [26] or conditional random fields [20]. Deep learning methods train models like transformers, whose learning and generalization capabilities surpass traditional approaches [34].

We adopt a deep learning approach using state-of-the-art transformers to classify sentence tokens into multiple classes.

6.4 Weakly Supervised Token Labeling

Existing research has mainly focused on joint learning of sentence and token labels. In zero-shot sequence labeling, token-level labels are inferred from a sentence classification model [23], enabling joint learning of sentence and token labels [10, 24]. A modified attention mechanism can also highlight key words as a proxy for token labeling [23, 24].

However, these approaches do not address scenarios where only coarse, instance-level annotations are available. In our task, domain experts provide annotations at the objective level rather than the token level. To leverage these partial annotations, we design a weakly supervised token labeling algorithm that converts objective-level labels into token-level supervision signals. Our method follows principles from weak supervision and partial label learning [3, 22], using coarse annotations as noisy signals to train sequence models effectively without requiring full token-level labels.

7 Conclusion

We propose a novel approach for automatically extracting key details from sustainability objectives using weak supervision. Our method tokenizes input text, applies a weakly supervised token-labeling algorithm to generate token-level labels from coarse annotations, and fine-tunes a transformer model for sequence labeling. Experimental results show that our approach outperforms state-of-the-art methods in sustainability objective detail extraction. We integrate our approach into GoalSpotter and demonstrate its post-deployment effectiveness in extracting structured details from diverse, heterogeneous data.

Future work will explore incorporating visual elements from sustainability reports, better handling of complex objectives, and enabling fuzzy matching to enhance extraction completeness and robustness.

References

[1] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L Opdahl, and Csaba Veres. 2020. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access* 8 (2020), 32862–32881.

[2] Marco Bronzini, Carlo Nicolini, Bruno Lepri, Andrea Passerini, and Jacopo Staiano. 2023. Glitter or gold? Deriving structured insights from sustainability reports via large language models. *arXiv preprint arXiv:2310.05628* (2023).

[3] Timothee Cour, Ben Sapp, and Ben Taskar. 2011. Learning from partial labels. *The Journal of Machine Learning Research* 12 (2011), 1501–1536.

[4] Jim Cowie and Wendy Lehnert. 1996. Information extraction. *Communications of the ACM* 39, 1 (1996), 80–91.

[5] Sebastião Vieira de Freitas Netto, Marcos Felipe Falcão Sobral, Ana Regina Bezerra Ribeiro, and Gleibson Robert da Luz Soares. 2020. Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe* 32, 1 (2020), 1–12.

[6] Vasiliki Gkatziaki, Symeon Papadopoulos, Richard Mills, Sotiris Diplaris, Ioannis Tsampoulatidis, and Ioannis Kompatsiaris. 2018. easIE: Easy-to-use information extraction for constructing CSR databases from the web. *ACM Transactions on Internet Technology (TOIT)* 18, 4 (2018), 1–21.

[7] Tushar Goel, Palak Jain, Ishan Verma, Lipika Dey, and Shubham Paliwal. 2020. Mining company sustainability reports to aid financial decision-making. In *Proceedings of the AAAI Workshop on Knowledge Discovery from Unstructured Data in Financial Services.*

[8] Matthias Gräuler and Frank Teuteberg. 2014. Greenwashing in online marketing–investigating trust-building factors influencing greenwashing detection. *Multikonferenz Wirtschaftsinformatik* (2014).

[9] Steven Katz, Yu Gu, and Lanxin Jiang. 2024. Information extraction from ESG reports using NLP: a ChatGPT comparison. *Available at SSRN 4836432* (2024).

[10] Canasai Kruengkrai, Thien Hai Nguyen, Sharifah Mahani Aljunied, and Lidong Bing. 2020. Improving low-resource named entity recognition using joint sentence and token labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*. 5898–5905.

[11] Dilek Küçük. 2022. Monitoring energy trends through automatic information extraction. *arXiv preprint arXiv:2201.01559* (2022).

[12] Guoliang Li, Dong Deng, and Jianhua Feng. 2011. Faerie: Efficient filtering algorithms for approximate dictionary-based entity extraction. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data.* 529–540.

[13] Alexandra Luccioni, Emily Baylor, and Nicolas Duchene. 2020. Analyzing sustainability reports using natural language processing. *arXiv preprint arXiv:2011.08073* (2020).

[14] Mohammad Mahdavi, Ramin Baghaei Mehr, and Tom Debus. 2024. Combat greenwashing with GoalSpotter: Automatic sustainability objective detection in heterogeneous reports. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM).*

[15] Mohammad Mahdavi, Amirhosein Farahmand, and Abolfazl Nadi. 2025. Fact-checking sustainability objectives using multimodal retrieval-augmented generation. In *2025 IEEE International Conference on Data Mining Workshops (ICDMW).*

[16] Zakariae Alami Merrouni, Bouchra Frikh, and Brahim Ouhbi. 2016. Automatic keyphrase extraction: An overview of the state of the art. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*. 306–313.

[17] Wayne Moodaley and Arnesh Telukdarie. 2023. Greenwashing, sustainability reporting, and artificial intelligence: A systematic literature review. *Sustainability* 15, 2 (2023), 1481.

[18] Zara Nasar, Syed Waqar Jaffry, and Muhammad Kamran Malik. 2021. Named entity recognition and relation extraction: State-of-the-art. *ACM Computing Surveys (CSUR)* 54, 1 (2021), 1–39.

[19] Oluyomi A Osobajo, Adekunle Oke, Ama Lawani, Temitope S Omotayo, Nkeiruka Ndubuka-McCallum, and Lovelin Obi. 2022. Providing a roadmap for future research agenda: A bibliometric literature review of sustainability performance reporting (SPR). *Sustainability* 14, 14 (2022), 8523.

[20] Fuchun Peng and Andrew McCallum. 2006. Information extraction from research papers using conditional random fields. *Information Processing & Management* 42, 4 (2006), 963–979.

[21] Marco Polignano, Nicola Bellantuono, Francesco Paolo Lagrasta, Sergio Caputo, Pierpaolo Pontrandolfo, and Giovanni Semeraro. 2022. An NLP approach for the analysis of global reporting initiative indexes from corporate sustainability reports. In *Proceedings of the First Computing Social Responsibility Workshop within the 13th Language Resources and Evaluation Conference.* 1–8.

[22] Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. 2017. Snorkel: Rapid training data creation with weak supervision. In *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*, Vol. 11. 269.

[23] Marek Rei and Anders Søgaard. 2018. Zero-shot sequence labeling: Transferring knowledge from sentences to tokens. *arXiv preprint arXiv:1805.02214* (2018).

[24] Marek Rei and Anders Søgaard. 2019. Jointly learning to label sentences and tokens. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6916–6923.

[25] Erik Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 142–147.

[26] Tobias Scheffer, Christian Decomain, and Stefan Wrobel. 2001. Active hidden markov models for information extraction. In *International Symposium on Intelligent Data Analysis*. 309–318.

[27] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. 1715–1725.

[28] Dominik Stammbach, Nicolas Webersinke, Julia Anna Bingler, Mathias Kraus, and Markus Leippold. 2022. A Dataset for detecting real-world environmental claims. *arXiv preprint arXiv:2209.00507* (2022).

[29] TerraChoice. 2010. The sins of greenwashing: Home and family edition. *Underwriters Laboratories* (2010).

[30] United Nations. 2023. Sustainable Development Goals. https://sdgs.un.org/goals. Accessed: February 20, 2023.

[31] Vinicius Woloszyn, Joseph Kobti, and Vera Schmitt. 2021. Towards automatic green claim detection. In *Forum for Information Retrieval Evaluation*. 28–34.

[32] Marco Wrzalik, Florian Faust, Simon Sieber, and Adrian Ulges. 2024. NetZeroFacts: Two-stage emission information extraction from company reports. In *Proceedings of the Joint Workshop of the 7th Financial Technology and Natural Language Processing, the 5th Knowledge Discovery from Unstructured Data in Financial Services, and the 4th Workshop on Economics and Natural Language Processing@ LREC-COLING 2024*. 70–84.

[33] Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access* 7 (2019), 173111–173137.

[34] Yang Yang, Zhilei Wu, Yuexiang Yang, Shuangshuang Lian, Fengjie Guo, and Zhiwei Wang. 2022. A survey of information extraction based on deep learning. *Applied Sciences* 12, 19 (2022), 9691.