Analysis of Global Reporting Initiative Sustainability Reports with NLP techniques

Sergio Caputo
University of Bari Aldo Moro
s.caputo34@studenti.uniba.it

Abstract

This paper aims to investigate and compare different Natural Language Processing (NLP) approaches to retrieve relevant information from sustainability reports in accordance with the Global Reporting Initiative framework. The proposed system allows to identify and locate references to the various sustainability topics discussed by the reports, to extract the contexts of each reference and to study if it has positive or negative sentiment. The output of the system has been then evaluated against a ground truth obtained through a manual annotation process on 134 reports. Experimental outcomes highlight the affordability of the approach for improving sustainability disclosures, accessibility, and transparency, thus empowering stakeholders to conduct further analysis and considerations.

1 Introduction

In recent decades corporate sustainability has become increasingly prominent in politics, management practice, reporting, and business science (Kolk, 2003 [2]; Barkemeyeret al., 2009 [3]). Customers, shareholders and investors ask for a greater transparency and regular disclosure of non-financial performance of companies.

Consumers and investors want to make informed choices and rational investment and that is the reason why they demand the disclosure of reliable information from enterprises (Epstein, 2008 [4]).

An organization can inform all interested parties about its economic, social and environmental activities by developing and issuing a sustainability report. CSR (Corporate Social Responsibility) reports are a voluntary business communication tool, which aim is to communicate the company attitudes towards assumptions of the CSR concept.

For the purpose of this paper sustainability reporting is considered as the practice of providing information to external and internal stakeholders on the economic, environmental and social results achieved by an organization. In literature and business practice most frequently used terminology for these kinds of practices are: sustainability reporting, corporate social responsibility reporting, non-financial disclosures.

Sustainability reporting is now a global standard. From year to year the number of reporting enterprises are growing. The European Union is the most active region in the world in terms of sustainability reporting. According to GRI statistics (GRI 2010), 45 % of published worldwide sustainability reports in 2010 year came from Europe.

2 Literature standards review

2.1 Global Reporting Initiative Standards

The Global Reporting Initiative Standards (GRI) represent global best practices for sustainability reporting of businesses, companies and institutions of any size anywhere in the world. The standards allow organizations to uniquely and uniformly measure their impact on planet Earth and make it public in a format that even non-experts in the field can understand.

It was originally a division of CERES (Coalition for Environmentally Responsible Economies) created to develop a sustainable accounting system that would allow companies to track their environ-

mental impact. This would make it easier for them to pursue goals within broader social responsibility. The GRI department was then recognized as an independent body in 2002, when UNEP (United Nations Environment Programme) shared its principles for member nations to follow.

Sustainability reporting based on the Standards provides information about an organization's positive or negative contribution to sustainable development and allows it to report on its economic, environmental and social impacts, thereby increasing transparency on their contribution to sustainable development.

In addition to reporting companies, the Standards are highly relevant to many stakeholders - including investors, policymakers, capital markets, and civil society.

The GRI Standards, which are modular and inter-correlated, are primarily designed to be used as a set, to prepare a sustainability report focused on material issues, their related impact and how they are managed. (Fig.1)
The GRI Standards are a modular system comprising three series of Standards:

1. **GRI Universal Standards**:
    - *GRI 101 Foundation* explains how to use the Standards and how to draft the report. It also specifies the principles – such as accuracy, balance, and verifiability – fundamental to good-quality reporting.
    - *GRI 102 General Disclosures* contains disclosures relating to details about an organization's structure and reporting practices; activities and workers; governance; strategy; policies; practices and stakeholder engagement. These give insight into the organization's profile and scale, and help in providing a context for understanding an organization's impacts.
    - *GRI 103 Material Topics* explains the steps by which an organization can determine the topics most relevant to its impacts, its material topics, and describes how the Sector Standards are used in this process. It also contains disclosures for reporting its list of material topics; the process by which the organization has determined its material topics and how it manages each topic.
2. **GRI Sector Standard**: intends to increase the quality, completeness, and consistency of the reports of the organizations. Standards are developed for 40 sectors, starting with those with the highest impact. The standards list topics that are likely to be relevant for most organizations in a given industry, and indicate the important information to report on these topics.
3. **GRI Topic Standards**: contain disclosures for providing information on topics. Examples include Standards on waste, occupational health and safety, and tax. Each Standard incorporates an overview of the topic and disclosures specific to the topic and how an organization manages its associated impacts.

Each Standard begins with a detailed explanation of how to use it. The companies may use all or part of some GRI Standards to report specific data. This can both be seen as the strength as well as the weakness of the GRI guidelines. In the years to come, the GRI has developed more and more, expanding the number of activities its principles address.

There exist three different sets of Topic Standards used in the referenced reports for this work, they cover respectively: Economy (GRI 200), Environment (GRI 300) and Social (GRI 400). Each of these Topic Standards is structured with a list of relevant disclosures about specific topics and each of these has fine grained informative to be reported.

These metrics gather accountability information, enabling businesses to identify potential risks and address them, possibly turning them into opportunities or strengths. In practice, the GRI Standards not only provide an opportunity for the business to change old polluting habits, but also analyze waste to reduce costs and increase efficiency in all production, storage and distribution processes. GRI's objective is to drive positive change and have a real impact on the social well-being of companies, keeping the focus on opportunities for better work for employees, more sustainability for the planet and the abolition, once and for all, of all forms of human exploitation.

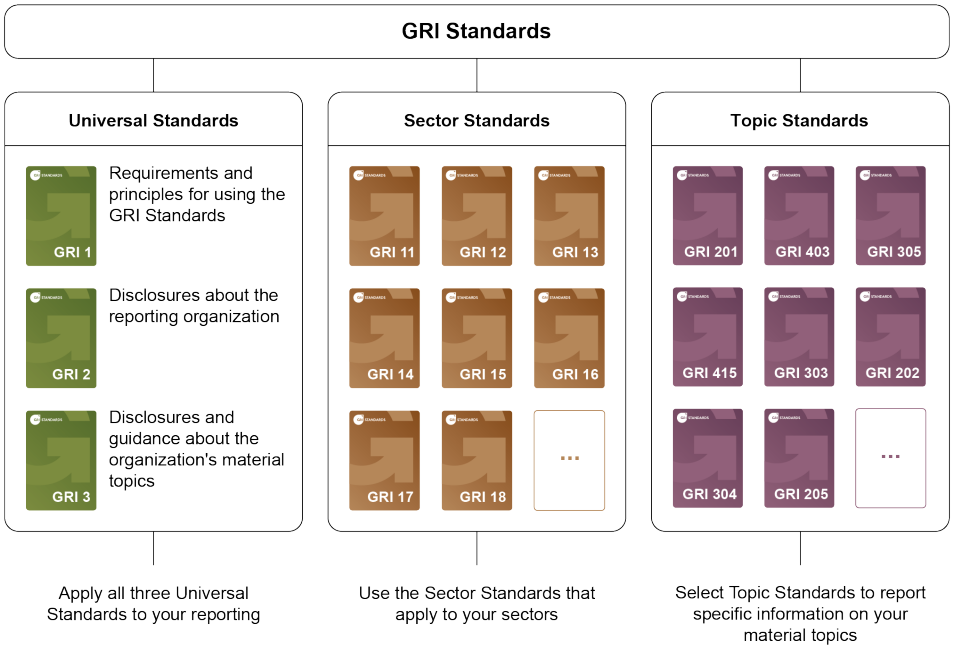

Figure 1: GRI Standard modules.

2.2 Non-financial Reporting Directive

Non-financial Reporting Directive (NFRD) is an amendment to the Accounting Directive (Directive 2013/34/EU) and was adopted in 2014. The disclosure of non-financial information is considered as vital for managing change towards a sustainable global economy by combining long-term profitability with social justice and environmental protection. The objective of the NFRD is therefore to raise the transparency of the social and environmental information provided by undertakings in all sectors to a similarly high level across all Member States and thus to improve the disclosure of non-financial information by certain large undertakings.

2.3 EU Corporate Sustainability Reporting Directive

The EU Corporate Sustainability Reporting Directive (CSRD), proposed on 21 April 2021, has a significantly extended scope than the Non-Financial Reporting Directive, applying to all large or listed companies operating in the EU. With a stated aim of bringing sustainability reporting on a par with financial reporting, it would help ensure both have equal weight and rigor.

It extends the reporting obligation to all large companies, all banks and all European insurance companies, whether quoted or unquoted, as well as all quoted companies, with the sole exception of quoted micro-companies. The reporting obligation is also extended to all groups, which will have to produce a consolidated sustainability report. In addition to listed micro-businesses, economic activities that are part of the consolidated reporting of the parent company, which is required to comply with the rules of European standards, are also excluded from the reporting obligation.

In numerical terms, applying the new inclusions and exclusions of the Directive, estimates predict that the 11,700 companies in Europe that are currently subject to the reporting obligation will increase to 49,000 economic activities.

3 GRI usage

A growing number of companies refers to the GRI as having inspired their reporting. Slightly over half (52.5%) of the Fortune Global 250 in the 2005 survey mentioned GRI in the report.
This does, however, not mean that companies follow the guidelines strictly, fully or consistently – they frequently take some components of the extensive set.

In 2005, 29% of the reports was specific about what parts of GRI were used and 6% declared to be in accordance with GRI. Companies sometimes also use the GRI guidelines to select the issues they include in their reports; this applied to 40% of the Fortune Global 250, and it turned out to be the most frequently mentioned tool to do this ('stakeholder consultation' in general was second with a score of 20%). Another way is which GRI turns out to play some role is in external verification –9% mentioned that the guidelines were part of this in one way or the other.

Studies within the available literature have pointed out that organizations have both positive and negative motivations to create sustainability reports. The positive motivations are linked to transparency and accountability, whereas the negative motivations tend to be linked to superficial aspects as enhancing an organization's image and decision-making direction sense, without substantive change. Different kind of problems have been found in the reporting model proposed by the GRI, it allows companies to report the positive facts only, while omitting the negative information. Therefore, legitimacy with no real transparency would be possible, allowing the report to be turned into a mere device to simulate sustainable positioning and to improve company's image. (Quilice al. 2018 [1])

4 Reporting Analysis

4.1 Research Goals

In this work, we were focused on the automatic analysis of textual documents concerning sustainability. In particular, we wanted to investigate the possibility of adopting NLP and IR techniques to be able to automatically extract relevant information for possible consultation and review by stakeholders. Specifically, it was considered that the preliminary analysis that could be done is to check whether specific sustainability topics or disclosures are discussed in the document. In this work, we focused on sustainability reports compliant with GRI Standards as the latter are by far the most widely adopted. This task, which might seem simple, is instead made complex by the heterogeneity of layouts and the writing style of sustainability documents.
The research questions we wanted to investigate were the following:

- **RQ1**: Is it possible to develop a robust and effective system for automatically search GRI topics from corporate sustainability reports?
- **RQ2**: How system performances are influenced by the granularity chosen for the keywords used while searching for GRI standards?
- **RQ3**: How the system performances are influenced by the tool used for the text extraction from PDF files?
- **RQ4**: Is it possible to use a sentiment analysis tool for evaluating if sustainability reports are discussing only positive aspects?

4.2 Dataset

In order to accomplish the proposed goal we created an ad hoc dataset of reports. The collection of PDF files was created collecting annual reports made publicly available by 134 Italian companies from 27 different sectors (e.g. waste management, automotive, agriculture) published by micro (2/134), small (2/134), medium (4/134) and large (126/134) organizations. The unbalanced representation with

respect to the drafting organizations size is due to the European norms, which state that non-financial disclosures are mandatory for large companies.

A large amount of companies publish environmental reports at least once per year, some other also quarterly depending on time-based objectives and goals to reach.

Companies are not forced to use a predefined layout design standard in drafting of reports: different designing choices are taken for characters style, sections design and disposition, graphs, images, number of pages, etc. Most of the reports are divided into 5 main sections, including management information, environment and climate change, environmental performance review, listing of verifiable environmental claims and green initiatives and declarations about environmental compliance.

In addition, also information about internal organization's structure, departments responsible, employees' roles in sustainability activities are provided in the reports.

In order to have a Ground Truth dataset, for each of the PDF reports, information about GRI disclosure presence were manually extracted, organized with respect to 118 GRI standards and collected in an Excel file.

4.3 OCR-based approach

Considering the format of files in the dataset, we had to convert each PDF report file into a set of images that could be processed by an OCR (Optical Character Recognition) tool in order to extract a textual representation of the pages inside the report. The extracted text was useful to perform a targeted search on GRI disclosures presence.

4.3.1 PDF to images conversion

First of all we had to make a suitable input for the OCR tool. For this purpose we used Poppler, a library for rendering PDF files, and examining or modifying their structure. It was used to convert PDF reports into a collection of images, in which each image corresponded to a page in the report.

4.3.2 Image to text conversion

After we performed the transformation of PDF files into collections of images, we adopted Tesseract, an OCR engine able to convert the text contained into an image, obtained with scans, pictures or photos, into understandable characters for a word processor. The results are usually very good as far as character recognition is concerned; however, the ability to maintain page layout, specifically for tables or columns, is lacking. Initially limited to ASCII characters, since March 2022 Tesseract supports UTF-8 characters and recognizes more than 100 languages

Given the textual representation of the reports, we were able to perform a key word search based on typical GRI's disclosures names in order to estimate the presence and absence of GRI Standards within the documents. Since we transformed each PDF page into an image, this allowed us also to detect where each of the GRI standard disclosure was located inside the extracted text from the documents analyzed.

4.3.3 Results Analysis

Considering the Ground Truth dataset with the real indications on presence or absence for each kind of disclosure, we were able to compute some metrics comparing these to the predicted ones by the simple key word search. We searched for 215 total keywords created by using the following two possible structures: "GRI <code >", "<code>". The element <code> is an integer number referring to GRI topics between 200 and 400 or their disclosures like 200-x, 300-x, and 400-x. Keywords like "GRI 302-4" or "GRI 203" or "306-4" are examples of of used search terms.
For each report we measured the:

- *Correct predictions for actually present disclosures (True Positives)*
- *Correct predictions for actually absent disclosures (True Negatives)*
- *Wrong predictions for actually present disclosures (False Negatives)*

- *Wrong predictions for actually absent disclosures (False Positives)*

Given these we could compute Precision and Recall measures in order to quantify the effectiveness of our method. Our approach (including GRI standards without specific disclosures, as "GRI 201", and fine grained ones) produced satisfying results considering Precision equal to 96% and Recall equal to 80% (Table 1).

Disclosures	**Real Present**	**Real Absent**
Predicted Present	4966	286
Predicted Absent	1368	9192

Table 1: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained ones).

In some cases we detected lower Recall values with respect to the average, this was caused by predicting absence of disclosures which are present instead. This phenomena showed up when we tried to search for key words with GRI names like "GRI 201" which did not refer to specific subcategory disclosures. In many reports it could happen that there would have been some references to generic GRI names just for mentioning purposes.

Since our goal was focused on detecting the presence of disclosures for reporting task, we also performed the same analysis only about specific subcategory disclosures. These kind of disclosures are characterized by names including the GRI standard name and a dash symbol before the numerical identifier for the distinct subcategory disclosure. Considering this new analysis we observed an improvement for the Recall value (89%) and an almost similar value for the Precision (95%) (Table 2).

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3496	242
Predicted Absent	511	7141

Table 2: Total predictions on individual reports for specific GRI subcategory disclosures presence.

Performance obtained was better than the one computed on general and specific disclosures, this is motivated by the fact that in many reports, specific disclosure references are much more commonly used for reporting purposes rather than general guideline references which may not be complete or detailed for the evaluation of reached goals by a company.

Some errors and misclassifications could be attributed to the OCR engine used which could not extract every word due to GRI words in images, sections, icons etc. or partial extractions of words difficult to detect and extract due to font type and color, background or layout design. Considering these problems we tried to adopt and compare the results obtained with an approach based on text extraction.

4.4 Text Extractor-based approach

In order to avoid errors in keywords matching caused by wrong detections and extractions made by the OCR engine, we used a simpler approach based on extracting text directly from the PDF files. For this purpose, PDFplumber was exploited. It is a Python open-source package whose objective is directed to parsing PDFs, analyzing PDF layouts and object positioning, and extracting text. For our case we used this library to extract text from PDF directly.

4.4.1 Results Analysis

As done for the OCR-based approach, we computed the Precision and Recall metrics for the set of disclosures that contained general and specific GRIs, and for the set limited to the specific standards.

Considering the set of general + specific GRI standards, results were quite similar to the ones obtained with the previous approach with 96% Precision and 81% Recall. (Table 3)

Disclosures	**Real Present**	**Real Absent**
Predicted Present	5047	275
Predicted Absent	1287	9203

Table 3: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained ones).

For the set of fine-grained only disclosures we observed the same results provided by the OCR-based approach on specific disclosures set (95% Precision, 90% Recall). A slight improvement was visible taking into account TP, FP, FN, TN. (Table 4)

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3542	231
Predicted Absent	465	7152

Table 4: Total predictions on individual reports for specific GRI subcategory disclosures presence.

Given the really similar results of the two approaches we investigated better the report cases for which Recall values were really low (<50%). In these circumstances we found out that for OCR-based approach we were having some problems related to low quality text extraction (caused by font type and color, background or layout design).

For the Text Extractor-based approach, low Recall cases were the ones related to reports with GRI standard references contained in images or graphical elements in the PDFs.

Taking into account the previous observations we decided to adopt a mixed approach based on both OCR and Text Extractor.

4.5 Mixed-based approach

For the Mixed-based approach we combined the results provided by the two methods previously described, in this way we could deal and cover a larger set of cases useful to detect GRI disclosures which were not considered by one of the two approaches.

More specifically, we combined the results of the methods such that cases not covered by OCR and covered instead by Text Extractor, or vice versa, could be considered for the disclosures detection.

For cases in which both methods could detect a GRI disclosure but indicating different pages where standard was found, we applied a rule of thumb: between two identified pages containing the GRI, only the one with greatest page number would have been the result considered by the Mixed approach. This decision was undertook since sustainability reports usually define GRI standards in the last pages of reports. This could also reduce errors that took into account presence of GRI disclosure in the opening pages which were used only for reference or for general indication.

4.5.1 Results Analysis

With the new approach and considering all the GRI types (general + fine-grained) we obtained 96% Precision and 84% Recall, very similar to previous approaches computed on general and specific standards. (Table 5)

Disclosures	**Real Present**	**Real Absent**
Predicted Present	5248	292
Predicted Absent	1086	9186

Table 5: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained ones).

These results were quite expected since we observed that general GRI disclosures do not provide additional information when sustainability reports are considered, usually only fine-grained standards are used to analyse the goals of companies.

Indeed, considering only fine-grained disclosures we reached 95% Precision and 94% Recall. (Table 6)

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3694	248
Predicted Absent	313	7135

Table 6: Total predictions on individual reports for specific GRI subcategory disclosures presence.

Errors obtained were relative to situations for which the approach detected the GRI standard disclosure presence, but it was not indicated in the Ground Truth (False Positive cases in the confusion matrix); or cases for which who annotated the Ground Truth considered disclosure as present even if there was not any reference to the GRI standard used, only a textual description about the topic of the GRI standard was present (False Negative cases in the confusion matrix).

5 Reporting Analysis considering Sentiment Analysis

Sustainability reporting should be considered as an impartial and transparent tool useful to explain to stakeholders, customers and lenders which are the sustainable goals, objectives and initiatives of companies. This mechanism could be exploited in a wrong way, companies might report the positive facts only, omitting the negative information, allowing companies to improve their own image, advertising projects and improving their sustainable position with respect to other companies.

In order to verify if this misuse is commonly applied, we conducted an analysis which took into account also the sentiment of the context where the GRI disclosures were found.

We would have expected inhomogeneous sentiments, since theoretically and ethically each company should make known information whether it is positive or negative, especially in the sustainability context.

Approaches explained above were reused considering also the sentiment information about contexts: when the standard's disclosure key word search produced a match inside the report, we took the match in consideration only if the context of the disclosure had positive or neutral sentiment. This modus operandi would have had us note the improved or worsened performance of each approach with respect to only positive or neutral disclosure's context.

For the Sentiment Analysis evaluation task we utilized two tools: TextBlob and Sent-It. The first one is a Python library for processing textual data with common Natural Language Processing (NLP) tasks, it was adopted to recognize the sentiment for contexts written in English.

Sent-It (Basile and Novielli, 2014 [5]) is a sentiment analysis tool indicated to identify the sentiment for texts written in Italian. It is a system based on a supervised machine learning approach. In particular, for training, three different kinds of features based on keywords and microblogging properties of tweets, on their representation in a distributional semantic model, and on a sentiment lexicon have been exploited. Data provided for training are annotated according to the subjectivity/objectivity of the content carried by the textual content. Moreover, each piece of text is categorized as positive, negative, or neutral. In our case, most of the disclosure's contexts were written in Italian, and we are able to obtain a score of polarity (i.e., positive, negative) and subjectivity/objectivity.

At the end of the analysis process, it was possible to export the results of each document in JSON format. It shows the reference context (if any), page number, polarity score, and subjectivity score for each GRI disclosure.

5.1 Sentiment Analysis study on contexts

The presence of certain negative words could definitely affect the sentiment processed using the Sent-It tool, in fact, in many cases, taking in consideration each of the contexts for which a GRI standard disclosure was present, we observed frequent negative words that affected the general sentiment computation. The words detected were about negative concepts, for example as "rischi", "corruzione", "malattia", "pericoloso" etc. These type of words could heavily influence the final sentiment prediction especially if these were found in sentences that referred to negation of negative concepts eg. "Non si registrano sanzioni", in this case even if the sentiment is neutral, the sentiment analysis tool detected negative sentiment. Similar situations were commonly found and a large number of contexts were classified as negative. An analysis on contexts defined as negative by Sent-It, suggested us to investigate more for some negative root words commonly used for reporting tasks. (Table 7)

Negative root word	**OCR-based contexts**	**Text Extractor-based contexts**
rischi	0,16	0,14
corruzione	0,14	0,14
rifiuti	0,08	0,09
discriminator	0,01	0,01
malatti	0,08	0,07
inquinant	0,01	0,01
spesa	0,01	0,01
emission	0,03	0,04
infortun	0,07	0,07
sanzion	0,05	0,05
incident	0,05	0,05
decess	0,03	0,03
pericolos	0,03	0,03
violazion	0,03	0,05
mort	0,02	0,02

Table 7: Percentage of contexts with negative root words and negative sentiment detected. For OCR-based approach negative contexts were 585, for Text-Extract based approach were 606.

In many cases it happened that these words were used in negations to underline the absence of problems or negative factors, but the sentiment analysis tool couldn't identify the actual neutral sentiment.

5.2 Results

Analyzing the results obtained adding sentiment information, we noticed performance was subjected to a reduction especially about Recall values for all the approaches. The reason this happened is related to the decrease in True Positive cases and to the increase of False Negative cases, as we could be expecting by only considering the contexts of positive/neutral standards disclosures. On average Recall values were 6,83% lower than the ones obtained by approaches without polarity consideration, this was sprung by sole consideration of disclosures with positive or at most neutral context, therefore the small Recall values reduction implied that companies documented negative aspects only in a few reports. This analysis aimed to highlight how companies used incorrectly and usually for promotional purposes a tool originally designed to be transparent and ethically proper.

5.2.1 OCR-based approach with Sentiment

Disclosures	**Real Present**	**Real Absent**
Predicted Present	4528	262
Predicted Absent	1806	9216

Table 8: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained one). (Precision = 96%, Recall = 74%).

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3148	219
Predicted Absent	859	7164

Table 9: Total predictions on individual reports for specific GRI subcategory disclosures presence. (Precision = 95%, Recall = 81%).

5.2.2 Text Extractor-based approach with Sentiment

Disclosures	**Real Present**	**Real Absent**
Predicted Present	4584	250
Predicted Absent	1750	9228

Table 10: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained one). (Precision = 96%, Recall = 74%).

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3180	206
Predicted Absent	827	7177

Table 11: Total predictions on individual reports for specific GRI subcategory disclosures presence. (Precision = 95%, Recall = 81%).

5.2.3 Mixed-based approach with Sentiment

Disclosures	**Real Present**	**Real Absent**
Predicted Present	4908	272
Predicted Absent	1426	9206

Table 12: Total predictions on individual reports for disclosures presence (including general GRI standards and fine grained one). (Precision = 96%, Recall = 80%).

Disclosures	**Real Present**	**Real Absent**
Predicted Present	3417	228
Predicted Absent	590	7155

Table 13: Total predictions on individual reports for specific GRI subcategory disclosures presence. (Precision = 95%, Recall = 87%).

6 Results evaluation

The results obtained from the different configurations let us answer to the previous research questions. Considering the results observable in Table 14, we can notice that F1 measure values are promising for all the configurations proposed. Specifically values change from the lowest value of 0.87210 obtained from the OCR-all-GRI configuration to 0.94365 found by performing the OCR-TE-sub-GRI configuration. These F1 measure values close to 1 underline how the proposed approaches are effective and reliable. Given these considerations we can positively answer to **RQ1**.

Comparing the configurations that considered both the general and the fine-grained disclosures, we observed some cases for which Recall values were lower with respect to the average, this problem was caused by the absence of matching with some general GRI standards used in the keyword search. Mismatching situations were caused by the manual annotator who considered general GRI disclosures present only because one of theirs fine-grained level disclosures was found in the report. For this reason we conducted a separated analysis for the sub-GRI level disclosures only and results emphasized the better behaviour than all the configurations that considered general and specific disclosures at the same time. These results support how the granularity chosen for the searching keywords is relevant for our goals (**RQ2**).

	Precision	Recall	F1
OCR -all-GRI	0.95579	0,80189	0.87210
OCR -sub-GRI	0.95103	0.89208	0.92061
TE -all-GRI	0.95752	0.80704	0.87586
TE -sub-GRI	0.95228	0.89616	0.92337
OCR-TE -all-GRI	0.95577	0.84152	0.89501
OCR-TE -sub-GRI	0.95014	0.93725	***0.94365***

Table 14: Results obtained from runs with different configurations

	Precision	Recall	F1
OCR-SA -all-GRI	0.95664	0,73529	0.83148
OCR-SA -sub-GRI	0.95236	0.81066	0.87582
TE-SA -all-GRI	0.95863	0.73673	0.83316
TE-SA -sub-GRI	0.95403	0.81050	0.87643
OCR-TE-SA -all-GRI	0.95642	0.79034	0.86548
OCR-TE-SA -sub-GRI	0.95106	0.87259	***0.91014***

Table 15: Results obtained from runs with Sentiment Analysis

Different configurations were tested in order to improve performance and quality of the extracted text from reports. The results obtained showed that using a text extractor succeeds in reducing some of the problems of OCR, particularly those in which the text was shown on colored backgrounds or in fonts that are difficult to interpret. On the contrary, the text contained in images is ignored. Indeed, we moved from an F1 measure value of 0.92061 for the OCR-based technique to 0.92337 for the one based on Text Extractor. Instead, the best performances were achieved when combining approaches. A GRI standard was considered identified if it was found in the text obtained by at least one of the two approaches. This process allowed us to obtain an F1 score of 0.94365, the highest among the results of our runs. These considerations allow us to provide a response to **RQ3**.

We conducted a study which considered also the sentiment information about the contexts within the disclosures were found, it pointed out how the performance of the approaches decreased if these were compared to the counterpart without sentiment analysis applied. Following a detailed analysis, it has been observed that the negative contexts identified were false positives. In fact, they were generated by the misclassification of the same by the sentiment analysis tool used. The presence of certain negative words could definitely affect the sentiment processed using the Sent-It tool. We detected that words such as "rischi", "corruzione", "malattia", "pericoloso" could heavily influence the final sentiment prediction especially if these were found in sentences that used negations. In sustainability reports, it is common that negative words are used in negations to strengthen a positive aspect.
This study highlighted how considering only the positive and neutral contexts slightly decreased the performance of the proposed methods, therefore it confirmed that companies tend to present only the positive aspects and goals achieved with respect to negative ones. The results obtained allow us to provide a clear answer to **RQ4**.

7 Conclusion

Sustainability reporting should be considered as an impartial, ethically proper, transparent and helpful tool to explain sustainable goals, objectives, and companies' activities to their stakeholders. Many companies base their sustainability reports on GRI standards or other reporting standards, but despite this, reports are poorly structured and thus complex to read and analyze. In this work we addressed the problem by proposing a system supporting the analysis of sustainability reports, specifically designed to identify the topics/disclosures discussed within GRI compliant reports. We proposed different Natural Language Processing and Information Retrieval approaches able to deal with closed format files, i.e. PDFs. The documents were transformed into a machine-readable textual format using OCR and Text Extraction tools. The text obtained here has been considered as the raw data over which to perform a keyword search operation. In particular, we considered keywords representative of the GRI topics and disclosures. The obtained results showed that the system proposed is extremely performant on the considered dataset, showing a score of F1 measure equal to 0.9436. It was also observed that the text extraction strategy from the PDF format could strongly impact on the obtained results, suggesting

a hybrid extraction mode to make up for the shortcomings of OCR and Text Extraction tools. A sentiment analysis study proved that the examined reports seemed to emphasize positive aspects rather than negative ones, which would contradict one of the GRI reporting principles (balance). A key future challenge is to enhance the system in terms of robustness, efficiency, effectiveness and flexibility.

References

[1] Quilice, TF, Cezarino, LO, Alves, MFR, Liboni, LB, Caldana, ACF. Positive and negative aspects of GRI reporting as perceived by Brazilian organizations. Environ Qual Manage. 2018; 27: 19–30. https://doi.org/10.1002/tqem.21543

[2] Kolk, A. (2003), Trends in sustainability reporting by the Fortune Global 250. Bus. Strat. Env., 12: 279-291. https://doi.org/10.1002/bse.370

[3] Barkemeyer, Ralf & Figge, Frank Hahn, Tobias Holt, Diane. (2009). What the Papers Say: Trends in Sustainability. A Comparative Analysis of 115 Leading National Newspapers Worldwide. Journal of Corporate Citizenship. 2009. 69-86. http://dx.doi.org/10.9774/GLEAF.4700.2009.sp.00009

[4] Epstein, M.J.: Making Sustainability Work. Best Practices in Managing and Measuring Corporate Social, Environmental, and Economic Impacts. Greenleaf Publishing, Sheffield (2008) http://dx.doi.org/10.4324/9781351276443

[5] Basile, P. and Novielli, N. (2014). Uniba at evalita 2014-sentipolc task predicting tweet sentiment polarity combining micro-blogging, lexicon and semantic features. UNIBA at EVALITA 2014-SENTIPOLC Task Predicting tweet sentiment polarity combining micro-blogging, lexicon and semantic features, pages 58–63