

## A Text-Mining Approach to the Evaluation of Sustainability Reporting Practices: Evidence from a Cross-Country Study

### Analiza znaczeniowa praktyk raportowania zrównoważonego rozwoju: perspektywa światowa

Ajay Kumar\*, Niladri Das\*\*

*Department of Management Studies, Indian Institute of Technology (ISM) Dhanbad,  
Jharkhand-826004, India*

*E-mails: \* ajay.dms16@gmail.com, ORCID: 0000-0003-2448-8322*

*\*\* niladri\_pnu2003@yahoo.co.in*

---

#### Abstract

This study examines the sustainability reports (SRs) of 200 firms in both developed and emerging economies in order to identify the words most frequently used in disclosing sustainability practices within the Triple Bottom Line (TBL) approach to reporting (which emphasizes economic, environmental, and social dimensions). Its aim is to evaluate these sustainability reports under the umbrella of the GRI framework. It adopts a semi-automated Text-Mining (TM) technique to evaluate the corporate SRs of select firms from the top ten economies by GDP at current prices. Based on the GRI Standards guidelines, a total of 208 keywords were identified for analysis. The disclosures were then awarded points based on the appearance of these keywords so that the appearance of one resulted in the awarding of a score of one; if a keyword did not appear then the report was scored a zero for that word. Furthermore, a wordcloud was also generated in order to better understand the inclination of reporting language towards various TBL reporting categories. This analysis of the SRs of 200 firms from the top ten economies of the world sheds light on the differences in reporting practices and priorities as they relate to various aspects of the GRI Standards guidelines. The results indicate that SR practices have grown rapidly in the last half decade of the period selected for study (2013-2017) as compared to the first half (2008-2012). Canada ranked highest for its disclosure practices in this analysis followed by the UK, Germany, US, Japan, France, Italy, Brazil, India, and China. This study found that all included countries improved their sustainability performance over the period 2008-2017.

**Key words:** sustainability reporting, GRI, CSR, top ten economies, text-mining, wordcloud.

#### Streszczenie

W niniejszym artykule przeanalizowano raporty dotyczące zrównoważonego rozwoju (SR) z 200 firm, zarówno w gospodarkach rozwiniętych, jak i wschodzących, w celu zidentyfikowania słów najczęściej używanych przy ujawnianiu praktyk zrównoważonego rozwoju w ramach podejścia do raportowania treple bottom line (TB, które kładzie nacisk na ekonomię, środowisko i wymiary społeczne. Celem jest ocena raportów dotyczących zrównoważonego rozwoju w ramach GRI. Przyjęto półautomatyczną technikę Text-Mining (TM) do oceny korporacyjnych praktyk na rzecz zrównoważonego rozwoju (SR) wybranych firm z dziesięciu największych gospodarek według PKB w cenach bieżących. W oparciu o wytyczne standardów GRI do analizy wytypowano łącznie 208 słów kluczowych. Przyznano im następnie punkty w oparciu o częstotliwość ich występowania, tak że pojawienie się jednorazowe skutkowało przyznaniem jednej punktacji; jeśli słowo kluczowe nie pojawiło się, raport był oceniany jako zero dla tego słowa. Ponadto utworzono chmurę słów, aby lepiej zrozumieć skłonność języka raportowania do różnych kategorii raportów TBL. Ta analiza rekomendacji 200 firm z dziesięciu największych gospodarek świata rzuca światło na różnice w praktykach i priorytetach raportowania, które odnoszą się do różnych aspektów wytycznych GRI. Wyniki wskazują, że praktyki zrównoważonego rozwoju (SR) gwałtownie wzrosły w ostatniej połowie dekady wybranej do badania (2013-2017), w porównaniu z pierwszą połową (2008-2012). W tej

analizie Kanada zajęła najwyższe miejsce pod względem praktyk ujawniania informacji, a następnie Wielka Brytania, Niemcy, Stany Zjednoczone, Japonia, Francja, Włochy, Brazylia, Indie i Chiny. Badanie wykazało, że wszystkie uwzględnione kraje poprawiły swoje wyniki w zakresie zrównoważonego rozwoju w latach 2008–2017.

**Słowa kluczowe:** raportowanie zrównoważonego rozwoju, GRI, CSR, 10 największych gospodarek, znaczenie tekstu, chmura słów

## Introduction

Climate change is a major barrier to sustainable development. The footprints of this change can be traced by analyzing the frequently observed unexpected natural disasters throughout the world, such as floods, droughts, earthquakes, etc. In addition, 2016 was classified as the *warmest year* according to statistics, indicating that global warming will be more brutal to the present civilization in the near future, if not effectively addressed (NASA, 2016). However, greenhouse gas (GHG) emissions are identified as one of the biggest causes of global warming and climate precariousness (IPCC, 1996). Due to major anthropogenic activities such as the combustion process of fossil fuels, carbon gasses are released into the environment. Therefore, carbon emissions, the most important GHGs, are responsible for stepping up the issue of atmospheric instability (Kaygusuz, 2009). About 90% of these anthropogenic activities are related to various business firms. Therefore, it is time for society to impose checks and balances on various business activities to decelerate the speed of environmental degradation and sustainability reports provide stakeholders with an effective tool to track their activities.

The first landmark report on sustainability was published in 1987, titled *Our Common Future*, it discussed the then current standards for sustainability reporting followed by major businesses (UNWCED, 1987). The report led to an increase in the responsibilities of corporations and led to the development of terms such as corporate social responsibility (CSR), sustainability reporting (SR), and corporate disclosing practices (CDPs). Interest in such issues grew and in just a few decades about 90-95% of large businesses began publishing their CSR reports or sustainability reports (SRs). The use of SRs has grown exponentially (Cho et al., 2012), resulting in the submission of millions of digital files and creating a virtual ocean of information on the internet.

Many organizations including the Sustainability Accounting Standards Board (SASB), The Prince's Accounting for Sustainability Project (A4S), World Business Council for Sustainable development (WBCSD), Account Ability (AA), Carbon Disclosure Project (CDP), Indian Centre for Corporate Social Responsibility (ICCSR), and the Global Report-

ing Initiative (GRI) are involved in attempts to improve the disclosure quality of SRs. These agencies have produced many guiding principles for SR practices, but the continuous revision of the GRI framework<sup>1</sup> and the simplicity of incorporating these guidelines into reporting practices has established GRI as the de facto standard globally (Spuerket et al., 2017). Most firms follow the GRI guidelines for filing their SRs in order to increase their credibility or reputation (Hedberg & von-Malmborg, 2003) and different stakeholders refer to them when investigating the content of disclosures (Searcy & Buslovich, 2014).

This study investigates the SRs of publicly listed companies in the top ten global economies as classified by the International Monetary Fund on the basis of GDP at current prices (World Bank, 2018). As compared to previous research in which content analysis was used to investigate such reports and which is limited to a few variables over a short span of time or a single country, this study is based on an investigation of corporate SRs produced by top ten economies between 2008-2017. This paper utilizes Text-Mining (TM) technique to help us gain a better understanding of international SR language in the dynamic business environment.

The remainder of the paper proceeds as follows. The next section 2 describes the review of literature and research objective while section 3 presents the methodological aspects of study. The results and discussion are shown in section 4 while the conclusion makes up section 5.

## Review of Literature

The first well-established definition of sustainable development published in the report of the Brundtland Commission in 1987 defines it as *meeting the needs of the present without compromising the ability of future generations to meet their own needs* (UNWCED, 1987). More often the terms CSR, CDP, and SR are used in an organizational context, yet they lack a concrete definition which is universally accepted (Linnenlueck & Griffiths, 2010). Despite this, SR practices can broadly be categorized into three types: economic, environmental, and social as mentioned in the Triple Bottom Line (TBL) approach promulgated by Elkington in 1994 (Kumar &

<sup>1</sup> The first version of the GRI framework G1 was published in 2000 followed by G2 in 2002, G3 in 2006, G3.1

in 2011, G4 in 2014, and the latest framework GRI Standards in 2016.

Das, 2018). The TBL concept focuses on the disclosure of facts and figures related to various parameters in the socio-economic and environmental areas. There has been a common perception that SR practices are primarily a western concern (Chapple & Moon, 2005) and that the firms in Asian countries often fall behind western firms on various parameters of SR (Baughn et al., 2007). But by publishing an increasing number of SRs (KPMG, 2016) and contributing to research in this area (Fifka, 2012), Asian countries such as India and China have begun contributing to the development of sustainable practices (Kumar & Das, 2018). The frequent financial shocks experienced by firms over the last few decades and an increasing awareness of sustainable development practices by civil society has also resulted in the rapid evolution of the concept of SRs globally (Beattie, 2014).

In order to interpret business understandings of how to cope with sustainability challenges some scholars have tried to decode the SR practices of different firms by studying their published SRs (Szekely & Brocke, 2017; Kuassi & Zinsou, 2018; Arena *et al.* 2018; Kumar & Das, 2018). Some have studied the frequency and quantity of information disclosed in the reporting (Kolk, 2004), while others apply purely qualitative content analysis to sketch out the firms' SR practices (Freundlieb & Teuteberg, 2013). A still developing branch of research in this field also adopts critical disclosure analysis (CDA) which is an interdisciplinary approach analyzing spoken and written text in order to investigate SR issues (Merkl-Davies & Koller, 2012). Some of these studies use predefined terms that limit the environment (i.e. ecological limit) by using context analysis (Bjørn et al., 2016).

Content analysis is generally understood to be a quantitative form of research, yet it significantly differs from the conventional approach of quantitative research as it involves the analysis of texts created by others and data that has not been generated by a pre-coded survey questionnaire method (Wilson & Rayson, 1993). Additionally, Artificial Intelligence (AI), Machine Learning (ML), or Text-Mining (TM) are now often incorporated into investigations focused on such tasks as drawing out key words from a full text (Deagan et al., 2004), tagging of semantic multilingual or cross-lingual words or metaphor linguistic analysis (Koller et al., 2008). Furthermore, TM (AI or ML) has also been used in research produced in such areas as biochemistry and biomedical sciences (McDonald et al., 2012), humanities and social sciences (Crawford Camiciottoli, 2010), and management sciences (Tregidga et al., 2007a).

Every year a large number of SRs are filed by various organizations and evaluating these together is a tedious job. Hence, there is a need for a smart and intelligent framework to evaluate these reports effortlessly and which would also have the capability

of extracting hidden information present in these reports. Prior efforts have been made to apply the ML technique to the evaluation of corporate SRs but it has been unable to uncover the types of content the researchers sought. Although, Modapothala & Issac (2009) in their study used this technique successfully to find the pattern line, but the scope of results was limited to a few variables only. Some scholars like Shirata et al., (2011) have developed a TM system to examine the financial reports of firms and concluded that it is easier to identify the potential threats from non-financial data compared to financial data. TM is also successfully implemented in areas of medical science for evaluating radiology reports (Prasad et al., 2010) and medical reports (Friedlin et al., 2011). TM has the potential to uncover latent information hidden in the different texts produced by firms. It also has the potential to save firms from future losses by revealing the true picture of political and consumer preferences. However, few scholars have explored this area through the application of text-mining (Liew et al., 2014, Modapothala et al., 2010), with most studies using a very limited number of reports for their analyses. TM uses a combination of various computer programming methods and tools to produce a text-corpus (i.e. words of bag) and to uncover discursive constructions used in the communication of information related to various parameters in corporate SRs (Jaworska & Nanda, 2016). Compared to other types of professional correspondence, corporate SRs remain an under-investigated area in management science.

Previous studies have either focused on mandatory or voluntary reporting practices using disclosures-analytics or linguistic analysis, which is conceptual in nature and often unable to highlight the dynamic nature of corporate SR practices. Qualitative disclosures analysis scholars consider the quantitative analysis of disclosures as deductive in nature (Tregidga et al., 2007b) while qualitative disclosures analyses are often criticized for being excessively subjective and lacking sound empirical grounding by quantitative scholars (Beattie, 2014). The primary research objective of this study is to generate a scorecard based on various parameters reported in SRs and the secondary objective is to uncover the inclination of language used in them by analyzing a wordcloud of the most frequently used words in order to increase our understanding of reporting practices broadly.

### Data Source and Research Methodology

The study adopts semi-automated-TM to evaluate the corporate SRs of select firms from the top ten economies. In this method, every text file itself is considered a concept vector and each concept vector contains a large number of terms used with a different frequency commonly referred to as the vector's dimensions. In other words, every unique term used

in the text file denotes a different dimension of a vector. In order to reduce the dimensionality of the vector and remove unwanted information, pre-processed text files have been used in this study (Crain et al., 2012). It utilizes the multiple keyword extracting technique to find information about the various GRI parameters. These keywords are unique terms that occur in the concept vector and altogether denote the corpus (Blei et al., 2003; Krestel et al., 2009). The steps used for the pre-processing of data are represented in Figure 1.

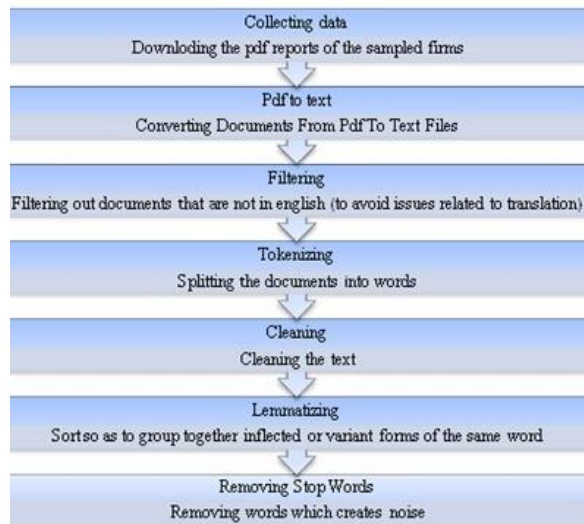


Figure 1. Steps of Pre-processing, own study

A sample of 200 firms (20 from each economy) has been selected from the different indices of select countries. The indices used are S&P 500 from the US, SSE 50 from China, NIKKEI 225 from Japan, DAX 30 from Germany, FTSE 100 from the UK, CAC 40 from France, SENSEX from India, FTSE MIB from Italy, BOVESPA 60 from Brazil, and S&P TSX 60 from Canada. Then, 20 firms were randomly selected from each of the selected indices. The firms for which sustainability reports, integrated reports, or CSR reports were found missing for any year (i.e. 2008-2017) were dropped from the sample and another firm was selected randomly from the affected index. 2008 was chosen as the starting point for this study because of the drastic increase in the filing of SRs by various organizations as per GRI guidelines. Before this year, very few SRs were filed (GRI, 2016). Furthermore, only SRs which were published in English were found appropriate for analysis

Ultimately, 2000 SRs published during 2008-2017 were included in the analysis. Various keywords from all 33 aspects of economic, environmental, and social reporting were extracted from the latest GRI guideline commonly known as GRI Standards. The list of keywords used is provided in annexure-1. A total of 208 keywords were identified for analysis. The appearance of each keyword leads to the awarding of a score of one for a report for that word and if

the word is not disclosed then the score is zero (Kumar & Das, 2018). For example, if all 208 keywords were present in a report then a firm can earn a maximum 208 score for that report. Similarly, the maximum score which can be earned by the firms is 2080 for the ten years of reporting. All semi-automated computer program codes were scripted in R using various packages such as tm, SnowballCC, RColorBrewer, ggplot2, wordcloud, biclust, Rcampdf, etc.

## Results and Discussions

During analysis, it was observed that the scores of firms for their SR practices in developed economies improved over the time period under study (i.e. 2008-2017). The results for SR practices in top ten economies have been classified into two categories: developed economies and emerging economies. The developed economies consist of US, Japan, Germany, UK, France, Italy, and Canada while emerging economies consist of China, India, and Brazil. The empirical results are shown in the following subsections. The annexure 1 represents the mean scores obtained by the respective countries, listed from the highest ranked to the lowest ranked country. In addition, at end of this section, figure 2 is the consolidated wordcloud diagram showing sustainable development in respected economies.

**Canada:** The number of Canadian firms filing SRs to GRI has been continuously increasing. According to the GRI database, only 47 firms had filed SRs in 2008 but it increased at CAGR of 10.27% and in 2017, a total of 125 firms had filed SRs (GRI, 2018). Canadian firms performed well in sustainability disclosures and scored the highest mean average score. The overall GRI scores of the 20 sampled firms jumped from 85.1 in 2008 to 142.21 in 2017 as shown in annexure 1.

The categorical investigation shows Canadian firms are more focused on environmental disclosures than social and economic disclosures. A possible reason for this may be the dominance of the Canadian mining industry throughout the world (MAC, 2017) and therefore in Canada, the orientation of its legal compliance is focused on preservation of the environment. GRI\_302, GRI\_305, and GRI\_306 were among the top three reported GRI parameters while GRI\_407, GRI\_206, GRI\_401 were least reported.

The linguistic analysis also confirms the pattern found during the categorical investigation as four out of ten words-emission, water, reuse, and pollution-were related to environmental aspects in the list of top ten most used words. The most frequent words used in reports filed by Canadian firms, as represented by the wordcloud in Figure 2a, are mine, emission, water, reuse, pollution, employees, program, support, service, and networks.

**United Kingdom (UK):** The number of UK firms filing SRs to the GRI has also been continuously increasing. According to the GRI database, only 68

firms had filed SRs in 2008 but it increased at CAGR of 13.53% and in 2017, a total of 242 firms had filed SRs.

After Canada, firms in the UK also performed well in sustainability disclosures and scored the second highest mean average score. The overall GRI scores of the 20 sampled UK firms jumped from 64.67 in 2008 to 134.61 in 2017 as shown in annexure 1.

The categorical investigation shows firms in the UK are more focused on social and environmental disclosures as compared to economic disclosures. GRI\_301, GRI\_304, and GRI\_306 were among the top three reported GRI parameters while GRI\_409, GRI\_205, GRI\_402 were least reported. The linguistic analysis of the SRs of UK firms is inclined towards social aspects of disclosure as seven out of ten words are related to society. The most frequent words used in reports filed by UK firms is represented in Figure 2b's wordcloud. It shows that the top ten words are people, community, ethics, positive, procurement, engagement, planet, business, performance, and scorecard.

**Germany:** The number of German firms filing SRs to the GRI has been increasing as well. According to the GRI database, only 48 firms had filed SRs in 2008 but it increased at CAGR of 11.68% and in 2017, a total of 145 firms had filed SRs. German firms are also doing well in sustainability disclosures and scored the third highest mean average score. The overall GRI scores of 20 sampled firms jumped from 62.12 in 2008 to 130.92 in 2017 as shown in annexure 1.

The categorical investigation shows German firms are more focused on environmental disclosures than social and economic ones. The possible reason for this may be Germany's stringent compliance measures focused on the preservation of the environment. GRI\_303, GRI\_305, and GRI\_302 were among the top three reported GRI parameters while GRI\_401, GRI\_203, GRI\_405 were least reported. The most frequent words used in reports filed by German firms represented in the Figure 2c wordcloud include employees, health, services, program, management, support, issues, company, reduce, and initiatives. The various policies encouraged German firms to implement green practices in their business operations as evidenced in this study by reporting on parameters such as GRI\_303, GRI\_305, and GRI\_302. The significance and current status of SR practices within Germany prove that the work of German firms toward achieving sustainable development (SD) has improved notably.

**United States (US):** The number of US firms filing SRs to the GRI also increased. According to the GRI database, only 133 firms had filed SRs in 2008 but it increased at CAGR of 29.28% and in 2017, a total of 347 firms had filed SRs. The US firms performed average in sustainability disclosures and have a mean score of 81.25. The overall GRI scores of 20

sampled firms jumped from 60.42 in 2008 to 125.39 in 2017 as shown in annexure 1.

As with Canadian firms, the categorical investigation shows that US firms have an approach more focused on environmental disclosures than social and economic ones. GRI\_302, GRI\_305, and GRI\_306 were among the top three reported GRI parameters while GRI\_407, GRI\_206, GRI\_401 were least reported. Linguistic analysis does not support the categorical analysis as the inclination of US firms leans more towards social issues. The most frequently used words in reports filed by US firms are represented in Figure 2d's wordcloud. The top ten words used by US firms are social, responsibility, principles, rights, planet, people, resources, employees, business, and services.

**Japan:** The number of Japanese firms filing SRs to GRI has been on the rise. According to the GRI database, only 183 firms had filed SRs in 2008 but it increased at CAGR of 1.15% and in 2017, a total of 205 firms had filed SRs. Japanese firms performed well in sustainability disclosures and scored the highest mean average score. The overall GRI scores of 20 sampled firms jumped from 57.86 in 2008 to 123.54 in 2017 as shown in annexure 1.

The categorical investigation shows Japanese firms have a more focused approach towards social disclosures as compared to environmental and economic disclosures. GRI\_403, GRI\_401, and GRI\_415 were among the top three reported GRI parameters while GRI\_308, GRI\_307, GRI\_304 were least reported. And the linguistic analysis of SRs shows that Japanese firms consider all three aspects of TBL reporting. The most frequently used words in reports filed by Japanese firms represented in the Figure 2e wordcloud include business, energy, management, corporate, production, plan, csr, income, committee, and hazardous.

**France:** The number of French firms filing SRs to GRI was also found to increase during the time period under study. According to the GRI database, only 32 firms had filed SRs in 2008 but it increased at CAGR of 19.17% and in 2017, a total of 185 firms had filed SRs. French firms did not perform well in sustainability disclosures and obtained the second lowest mean average score in the group of developed countries. In annexure 1 overall GRI scores of 20 sampled firms jumped from 57.01 in 2008 to 119.86 in 2017.

The categorical investigation shows French firms are more focused on environmental disclosures as compared to social and economic disclosures. While GRI\_302, GRI\_305, and GRI\_306 were among the top three reported GRI parameters, GRI\_407, GRI\_206, GRI\_401 were least reported. The linguistic analysis suggests that French firms have adopted a balanced approach to sustainability as the most frequently used words represent all three aspects of the TBL approach. Those words are found in the Figure





Figure 2a: Wordcloud Depicting Sustainable Development in Canada

Figure 2b: Wordcloud Depicting Sustainable Development in UK

Figure 2c: Wordcloud Depicting Sustainable Development in Germany

Figure 2d: Wordcloud Depicting Sustainable Development in US



Figure 2e: Wordcloud Depicting Sustainable Development in Japan

Figure 2f: Wordcloud Depicting Sustainable Development in France

Figure 2g: Wordcloud Depicting Sustainable Development in Italy

Figure 2h: Wordcloud Depicting Sustainable Development in Brazil



Figure 2i: Wordcloud Depicting Sustainable Development in India

Figure 2j: Wordcloud Depicting Sustainable Development in China

Figure 2. Wordclouds Depicting Sustainable Development in top ten economies, own's work

2f wordcloud and include million, water, emissions, management, performance, sustainability, co2, data, group, and solutions.

**Italy:** The number of Italian firms filing SRs to the GRI has been continuously increasing. According to the GRI database, only 48 firms had filed SRs in 2008 but it increased at CAGR of 8.44% and in 2017, a total of 108 firms had filed SRs.

Italian firms also performed below average in sustainability disclosures and scored the lowest mean average score of the developed countries examined in this study. The total overall GRI scores of 20 sampled firms jumped from 55.31 in 2008 to 108.79 in 2017 as shown in annexure 1.

The categorical investigation shows Italian firms are more focused on social disclosures than environmental or economic ones. GRI\_303, GRI\_405, and GRI\_406 were among the top three reported GRI parameters while GRI\_206, GRI\_403, GRI\_412 were least reported. A possible reason for this may be isomorphic pressure from civil society throughout the world. The linguistic analysis shows that Italian

companies give equal weight to all aspects of the TBL approach to reporting. The most frequently used words in the reports of Italian firms are shown in the Figure 2g wordcloud. These include performance, management, employees, production, environmental, indicators, emissions, company, local, and sustainability.

**Brazil:** The number of Brazilian firms filing SRs to GRI has been increasing. According to the GRI database, only 87 firms had filed SRs in 2008 but it increased at CAGR of 10.77% and in 2017, a total of 242 firms had filed SRs. When compared to the firms of developed economies, Brazilian firms did not perform well; however, among emerging economies Brazilian firms performed the best, obtaining the highest mean average score. The total overall GRI scores of 20 sampled firms jumped from 36.59 in 2008 to 79.29 in 2017 as shown in annexure 1.

The categorical investigation shows Brazilian firms have an approach focused on social disclosures as compared to environmental and economic disclosures. GRI\_402, GRI\_405, and GRI\_406 were

among the top three reported GRI parameters, while GRI\_407, GRI\_206, GRI\_401 were least reported. Linguistic analysis also supports the categorical analysis as five out of ten words are related to the environment. Figure 2h shows the most frequently used words in reports filed by Brazilian firms which include year, ensure, use, area, sea, air, social, environment, risk, and emission.

**India:** The number of Indian firms filing SRs to GRI has been growing. According to the GRI database, only 24 firms had filed SRs in 2008 but it increased at CAGR of 29.28% and in 2017, a total of 313 firms had filed SRs. Indian firms have shown the average result in the scoring of SR practices. The total overall GRI scores of 20 sampled firms jumped from 34.89 in 2008 to 73.76 in 2017 as shown in annexure 1.

The categorical investigation shows that GRI\_302, GRI\_305, and GRI\_306 were among the top three reported GRI parameters while GRI\_407, GRI\_206, GRI\_401 were least reported. The above results are also supported by the linguistic analysis of the SRs of Indian firms. As shown in the Figure 2i wordcloud, Indian firms most often used words related to the environmental category. The most frequently used words include water, business, waste, energy, development, emissions, performance, indian, sustainable, and products. The inclination of Indian firms towards the environmental aspects of reporting is most likely due to mandated reporting overseen by regulatory authorities. Traditionally the Indian approach to sustainability was cultivated through benevolence towards society but these days a western approach (i.e. sustainability leads to improved profitability in the long run) is increasingly adopted as Indian businesses more fully participate in the global economy (Balasubramanian et al., 2005).

**China:** The number of Chinese firms filing SRs to the GRI has grown during the period under study. According to the GRI database, only 28 firms had filed SRs in 2008 but it increased at CAGR of 40.05% and in 2017, a total of 813 firms had filed SRs. Though there was an increase in reporting, Chinese firms did not perform well in sustainability disclosures and scored the lowest mean average score. The total overall GRI scores of 20 sampled firms jumped from 30.63 in 2008 to 70.07 in 2017 as shown in annexure 1.

The categorical investigation shows Chinese firms have a focus on the social aspects of disclosures rather than environmental and economic aspects. GRI\_301, GRI\_206, GRI\_305 were among the top three reported GRI parameters while GRI\_302, GRI\_306, and GRI\_412 were least reported. The linguistic analysis of the SRs of Chinese firms mirrors the findings of the categorical analysis, showing an inclination towards language related to the social aspect of SRs. The most frequently used words in Chinese SRs are represented in the Figure 2j wordcloud and include csr, service, company, system, urban,

area, ensure, safe, access, and key. Many aspects of the government's reporting policies in China are voluntary and indicative (Marquis and Qian, 2014). These SR policies place more of an emphasis on environmental issues (i.e. GRI\_301, GRI\_305) than social issues, especially human-rights (i.e. GRI\_412), as indicated in the categorical analysis.

## Conclusion

This study has shown how firms in different nations utilize different SR practices as well as identifies differences in the disclosure of various aspects of the GRI Standards guidelines. By utilizing TM, this study was able to identify the words most often used in the writing of SRs, providing insight into what issues both companies and countries find most important. While this study identified differences across countries, it also showed that all countries have improved their sustainability performance over time (i.e. 2018-2017).

This study also ranked the SR practices of sampled economies after classifying them into developed economies and emerging economies. Canada ranked highest among the developed economies, followed by the UK, Germany, US, Japan, France, and Italy. In the emerging economies category, Brazil led the way followed by India and China. China is a particularly interesting case as Chinese firms received the lowest ranking in SR scoring, but China simultaneously had the highest CAGR, 40.05%. This would seem to suggest that Chinese firms are concerned with sustainability. A general observation across all countries is that reporting related to economic issues lags behind reporting on environmental or social issues. One possible reason for this may be the publication of separate financial statements in the form of annual reports as well as an understanding among firm management that sustainability reports are mostly for socio-environmental purposes.

There are some limitations to this study. It is focused on the SR practices of the largest countries and ignores the contributions of smaller countries on this issue. Considering them together might lead to a fuller understanding of global sustainability reporting practices. Another limitation is this study's focus on only the 20 largest firms in the countries examined. This opens up the possibility that the practices of medium- and small-sized enterprises might change the country rankings obtained in this study.

## References

1. ARENA C., LIONG R., VOURVACHIS P., 2018, Carrot or stick: CSR disclosures by Southeast Asian companies, in: *Sustainability Accounting, Management and Policy Journal*, 9(4), p. 422-454.
2. BALASUBRAMANIAM N.K., KIMBER D., SIEMENSMA F., 2005, Emerging opportunities or traditions reinforced? An analysis of the attitudes towards CSR, and trends of thinking about CSR India, in: *Journal of Corporate Citizenship*, 17, p. 79-92.

Annexure 1: GRI Ranked scorecard of Top Ten Economies, own work

Countries	Scores	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Avg GRI Score	Rank
Canada	Economic Scores	17.98	18.3	18.92	21.28	23.95	24.1	26.76	27.51	30.27	33.52	111.00	1
	Environment Scores	29.99	29.95	31.66	32.64	34.82	35.98	36.2	38.35	39.64	40.45		
	Social Scores	37.13	39.44	43.42	46.84	48.37	51.3	57.16	59.26	66.58	68.24		
	Total GRI Scores	85.1	87.69	94	100.76	107.14	111.38	120.12	125.12	136.49	142.21		
UK	Economic Scores	13.66	13.9	14.37	16.17	18.2	18.31	20.33	20.9	23	30.98	87.01	2
	Environment Scores	22.79	22.76	24.06	24.8	26.46	27.34	27.51	29.14	30.12	37.91		
	Social Scores	28.21	29.97	32.99	35.59	36.76	38.98	43.44	45.03	50.6	65.7		
	Total GRI Scores	64.67	66.64	71.44	76.57	81.42	84.64	91.29	95.09	103.73	134.61		
Germany	Economic Scores	13.12	13.35	13.81	15.53	17.48	17.59	19.53	20.08	22.09	29.75	83.74	3
	Environment Scores	21.89	21.86	23.11	23.82	25.41	26.26	26.42	27.99	28.93	36.68		
	Social Scores	27.1	28.79	31.69	34.19	35.31	37.44	41.72	43.25	48.6	64.47		
	Total GRI Scores	62.12	64.01	68.62	73.55	78.21	81.3	87.68	91.33	99.63	130.92		
US	Economic Scores	12.76	12.99	13.43	15.1	17	17.11	18.99	19.53	21.49	27.91	81.25	4
	Environment Scores	21.29	21.26	22.47	23.17	24.72	25.54	25.7	27.22	28.14	34.84		
	Social Scores	26.36	28	30.82	33.25	34.34	36.42	40.58	42.07	47.27	62.63		
	Total GRI Scores	60.42	62.25	66.74	71.53	76.06	79.07	85.28	88.83	96.9	125.39		
Japan	Economic Scores	12.22	12.44	12.86	14.47	16.28	16.38	18.19	18.7	20.58	27.29	78.16	5
	Environment Scores	20.39	20.36	21.52	22.19	23.67	24.46	24.61	26.07	26.95	34.22		
	Social Scores	25.24	26.81	29.52	31.85	32.89	34.88	38.86	40.29	45.27	62.01		
	Total GRI Scores	57.86	59.62	63.92	68.51	72.85	75.73	81.68	85.08	92.81	123.54		
France	Economic Scores	12.04	12.26	12.67	14.25	16.04	16.14	17.92	18.43	20.28	26.07	76.83	6
	Environment Scores	20.09	20.06	21.21	21.86	23.32	24.1	24.25	25.69	26.55	33		
	Social Scores	24.87	26.42	29.09	31.38	32.4	34.37	38.29	39.7	44.6	60.79		
	Total GRI Scores	57.01	58.75	62.98	67.5	71.78	74.62	80.48	83.83	91.44	119.86		
Italy	Economic Scores	11.68	11.89	12.29	13.83	15.56	15.66	17.39	17.88	19.67	22.38	73.78	7
	Environment Scores	19.49	19.46	20.57	21.21	22.63	23.38	23.53	24.92	25.76	29.31		
	Social Scores	24.13	25.63	28.22	30.44	31.44	33.34	37.15	38.51	43.27	57.1		
	Total GRI Scores	55.31	56.99	61.1	65.49	69.64	72.39	78.07	81.32	88.71	108.79		
Brazil	Economic Scores	7.73	7.86	8.13	9.15	10.29	10.36	11.5	11.82	13.01	12.54	49.54	8
	Environment Scores	12.89	12.87	13.61	14.03	14.97	15.47	15.56	16.49	17.04	19.47		
	Social Scores	15.96	16.95	18.67	20.14	20.79	22.05	24.57	25.48	28.62	47.26		
	Total GRI Scores	36.59	37.7	40.42	43.32	46.07	47.89	51.65	53.8	58.69	79.29		
India	Economic Scores	7.37	7.5	7.75	8.72	9.81	9.88	10.97	11.27	12.41	10.7	47.05	9
	Environment Scores	12.29	12.27	12.98	13.38	14.27	14.75	14.84	15.72	16.25	17.63		
	Social Scores	15.22	16.17	17.8	19.2	19.83	21.03	23.43	24.29	27.29	45.42		
	Total GRI Scores	34.89	35.95	38.54	41.31	43.92	45.66	49.24	51.29	55.96	73.76		
China	Economic Scores	6.47	6.58	6.81	7.66	8.62	8.67	9.63	9.9	10.89	9.47	41.84	10
	Environment Scores	10.79	10.78	11.39	11.75	12.53	12.95	13.03	13.8	14.27	16.4		
	Social Scores	13.36	14.19	15.63	16.86	17.41	18.46	20.57	21.33	23.96	44.19		
	Total GRI Scores	30.63	31.56	33.84	36.27	38.57	40.09	43.24	45.04	49.13	70.07		



3. BAUGHN C., BODIE N., MCINTOSH J., 2007, Corporate social and environmental responsibility in Asian countries and other geographical regions, in: *Corporate Social Responsibility and Environmental Management*, 14(4), p. 189-205.
4. BEATTIE V., 2014, Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework, in: *The British Accounting Review*, 46, p. 111-34.
5. BJØRN A., BEY N., GEORG S., RØPKE I., HAUSCHILD M.Z., 2017, Is Earth recognized as a finite system in corporate responsibility reporting?, in: *Journal of Cleaner Production*, 163, p. 106-117.
6. BLEI D.M., NG A.Y., JORDAN M.I., 2003, Latent dirichlet allocation, in: *Journal of machine Learning research*, 3, p. 993-1022.
7. CHAPPLE W., MOON J., 2005, Corporate social responsibility (CSR) in Asia: A seven-country study of CSR web site reporting, in: *Business & society*, 44(4), p. 415-441.
8. CHO C.H., MICHELON G., PATTEN D.M., 2012, Enhancement and obfuscation through the use of graphs in sustainability reports: An international comparison, in: *Sustainability Accounting, Management and Policy Journal*, 3(1), p. 74-88.
9. CRAIN S.P., ZHOU K., YANG S.H., ZHA H., 2012, Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond, in: *Mining text data*, Springer, Boston, MA, p. 129-161.
10. CRAWFORD CAMICIOTTOLI B., 2010, Earnings calls: Exploring an emerging financial reporting genre, in: *Discourse & Communication*, 4(4), p. 343-359.
11. DEEGAN M., SHORT H., ARCHER D., BAKER P., MCENERY T., RAYSON P., 2004, Computational linguistics meets metadata, or the automatic extraction of key words from full text content, in: *RLG Diginews*, 8(2).
12. DFAIT 2009, *Mining materials*, <https://www.nrcan.gc.ca/mining-materials/publications/8776> (12.05.2018).
13. ELKINGTON J., 1994, Towards the sustainable corporation: Win-win-win business strategies for sustainable development, in: *California management review*, 36(2), p. 90-100.
14. FIFKA M., 2012, The development and state of research on social and environmental reporting in global comparison, in: *Journal für Betriebswirtschaft*, 62(1), p. 45-84.
15. FREUNDLIEB M., TEUTEBERG F., 2013, Corporate social responsibility reporting-a transnational analysis of online corporate social responsibility reports by market-listed companies: contents and their evolution, in: *International Journal of Innovation and Sustainable Development*, 7(1), p. 1-26.
16. FRIEDLIN J., MAHOUI M., JONES J., JAMIESON P., 2011, Knowledge discovery and data mining of free text radiology reports, in: *Healthcare Informatics, Imaging and Systems Biology (HISB)*, 2011 First IEEE International Conference on, p. 89-96.
17. GRI, 2016, *20 years on – is sustainability reporting actually making a difference*, <https://www.globalreporting.org/information/news-and-press-center/Pages/20-years-on-Is-sustainability-reporting-actually-making-a-difference.aspx> (11.05. 2018).
18. GRI, 2018, *Global reporting*, <http://database.globalreporting.org/search/> (10.05.2018).
19. HEDBERG C.J., VON MALMBORG F., 2003, The global reporting initiative and corporate sustainability reporting in Swedish companies, in: *Corporate social responsibility and environmental management*, 10(3), p. 153-164.
20. IPCC 1996, *Second assessment on climate change*, International panel on climate change, Cambridge University Press: Cambridge.
21. JAWORSKA S., NANDA A., 2016, Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility, in: *Applied Linguistics*, 39(3), p. 373-399.
22. KAYGUSUZ K., 2009, Energy and environment issues relating to greenhouse gas emissions for sustainable development in Turkey, in: *Renewable and Sustainable Energy Reviews*, 13, p. 253-270.
23. KOLK A., 2004, A decade of sustainability reporting: developments and significance, in: *International Journal of Environment and Sustainable Development*, 3(1), p. 51-64.
24. KOLLER V., HARDIE A., RAYSON P., SEMINO E., 2008, Using a semantic annotation tool for the analysis of metaphor in discourse, in: *Metaphorik.de*, 15(1), p. 141-160.
25. KPMG 2016, *CSR survey*, <https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/02/CSR-Survey-2016.pdf> (05.05. 2018).
26. KRESTEL R., FANKHAUSER P., NEJDL W., 2009, Latent dirichlet allocation for tag recommendation, in: *Proceedings of the third ACM conference on Recommender systems*, p. 61-68.
27. KUASSI M., ZINSOU CH., 2018, Integrated or non-integrated reports: French listed companies at a crossroads?, in: *Sustainability Accounting, Management and Policy Journal*, 9(3), p. 253-288, DOI: 10.1108/SAMPJ-09-2017-0114.
28. KUMAR A., DAS N., 2018, Sustainability Reporting Practices in Emerging Economies: A Cross-Country Study of BRICS Nations, in: *Problemy Ekorożwoju/ Problems of Sustainable Development*, 13(2), p. 17-25.
29. LIEW W. T., ADHITYA A., SRINIVASAN R., 2014, Sustainability trends in the process industries: A text mining-based analysis, in: *Computers in Industry*, 65(3), p. 393-400, DOI: 10.1016/j.compind.2014.01.004
30. LINNENLUECKE M.K., GRIFFITHS A., 2010, Corporate sustainability and organizational culture, in: *Journal of world business*, 45(4), p. 357-366.
31. MAC, 2017, Facts and figures 2017, <http://mining.ca/sites/default/files/documents/Facts-and-Figures-2017.pdf> (20.05.2018).
32. McDONALD D., MCNICOLL I., WEIR G., REIMER T., REDFEARN J., JACOBS N., BRUCE R., 2012, The value and benefits of text mining, in: *JISC Digital infrastructure*.
33. MERKL-DAVIES D.M., KOLLER V., 2012, Metaphoring' people out of this world: A Critical Discourse Analysis of a chairman's statement of a UK defence firm in: *Accounting Forum*, 36(3), p. 178-193.
34. MODAPOTHALA J.R., ISSAC B., 2009, Evaluation of corporate environmental reports using data mining approach, in: *2009 International Conference on Com-*

- puter Engineering and Technology, IEEE. p. 543-547.
35. MODAPOTHALA J.R., ISSAC B., JAYAMANI E., 2010, Appraising the corporate sustainability reports-text mining and multi-discriminatory analysis, in: *Innovations in Computing Sciences and Software Engineering*, Springer, Dordrecht, p. 489-494.
  36. NASA, 2016, *NASA NOAA data show 2016 warmest year on record globally*, <https://www.nasa.gov/pres-s-release/nasa-noaa-data-show-2016-warmest-year-on-record-globally> (20.01.2018).
  37. PRASAD A.K., RAMAKRISHNA S., KUMAR D.S., RANI B.P., 2010, Extraction of radiology reports using text mining, in: *International Journal on Computer Science and Engineering*, 02(05), p. 1558-1562.
  38. SEARCY C., BUSLOVICH R., 2014, Corporate perspectives on the development and use of sustainability reports, in: *Journal of business ethics*, 121(2), p. 149-169.
  39. SHIRATA C.Y., TAKEUCHI H., OGINO S., WATANABE H., 2011, Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining, in: *Journal of Emerging Technologies in Accounting*, 8(1), p. 31-44, DOI: 10.2308/jeta-10182.
  40. SPUERK S., DROBE M., AND LOTTERMOSER B.G., 2017, Evaluating resource efficiency at major copper mines, *Minerals Engineering*, 107, p. 27-33.
  41. SZÉKELY N., AND VOM BROCKE J., 2017, What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique, in: *PloS one*, 12(4).
  42. TREGIDGA H., MILNE M., KEARINS K., 2007b, The role of discourse in bridging the text and context of corporate social and environmental reporting, in: *Asia-Pacific Interdisciplinary Research into Accounting Conference*, Auckland, NZ, 7-9 July.
  43. TREGIDGA H., MILNE M.J., KEARINS K., 2007b, The role of discourse in bridging the text and context of corporate social and environmental reporting, in: *Department of Accountancy, Finance and Information Systems*, University of Canterbury.
  44. UNWCED, 1987, *Our Common Future*, <http://www.un-documents.net/our-common-future.pdf> (01.05.2018).
  - WEF, 2018, *The world's biggest economies in 2018*, <https://www.weforum.org/agenda/2018/04/the-world-s-biggest-economies-in-2018/> (30.04.2018).
  45. WILSON A., & RAYSON P., 1993, Automatic Content Analysis of Spoken Discourse, in: *Corpus Based Computational Linguistics*, eds. Souter C., Atwell E., Rodopi, Amsterdam, p. 215-226.
  46. WORLD BANK, 2018, *GDP*, [https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2016&start=1960&view=chart&year\\_high\\_desc=true](https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2016&start=1960&view=chart&year_high_desc=true) (22.04.2018).