CSR IN THE CZECH REPUBLIC THROUGH THE LENS OF HIGH DIMENSIONAL DATA ANALYSIS

Xhesilda Vogli

Department of Management, Faculty of Economics and Management, Czech University of Life Sciences Prague, 165 21 Prague, Czech Republic

Abstract

This study analyses the websites of one thousand worldwide companies that operate currently in the Czech Republic focusing on headings that relate to corporate social responsibility. All data was collected from websites of companies in the Czech Republic where common language was analyzed with a machine learning algorithm converting the result into a proper dataset for graph theory using high dimension data analysis. The purpose of this research paper is to identify the common terminology used for CSR and determine if factors such as company size or industry affect the interpretation of CSR. The paper provides visibility of the current understanding of CSR in companies that operate in the Czech Republic. The findings of this work can be used to develop further frameworks and management strategies in order to better communicate CSR initiatives to stakeholders being those external or internal.

Key words: *Corporate Social Responsibility, environment, sustainability, diversity, management, framework, high dimension data analysis, UNMAP*

1. INTRODUCTION

The financial crisis that began in the summer of 2007 in the United States that still characterizes the global economy has triggered also a debate regarding the ratio between ethic and business making us witnesses of the radical change between private business and civil society. [1] Although the origin of Corporate Social Responsibility has roots in the ’20, the topic has been referred under many names such as sustainability, voluntary initiatives, philanthropy or even just social responsibility. Corporates have been called upon modifying their strategies beyond financial aspects and incorporate into their actions also their impact in civil society considering social and environmental impact. According to Forbes, 42% of how people see a company is based on their perceptions of the firm’s corporate social responsibility practices.[2] As an example, Microsoft has been carbon neutral across the world since 2012 and it’s committed to be carbon negative by 2030. [3] SAP has incorporated CSR into their vision and mission statement and runs initiatives like ‘Meet and Code’ from which in 2018 benefited 34,000 teachers and engaged more than 2.8 million students.[4] Whereas other companies like Upfield are working to build a sustainable business starting from supply chains, operations and interactions with the civil world. They are a good example of responsible sourcing because they rely on ingredients such as palm, olive, or sunflower oil to make their products. All their products don’t contribute in any deforestation, destruction of habitats or diversity. [5] Different authors have tried to identify the evolution of the CSR concept throughout the years.

Milton Friedman, in his book Capitalism and Society defines corporate social responsibility a “*fundamentally subversive doctrine*,” arguing instead that the goal of business should be merely to amplify profits. Although in the following definition he recognizes on top of what has been embodied by law, companies must comfort in an ethical and moral basic rule of the society: “*In a free-enterprise, private-property system, a corporate executive is an employee of the owners of the business. He has direct responsibility to his employers. That responsibility is to conduct the business in accordance with their desires, which generally will be to make as much money as possible while conforming to the basic rules of the society, both those embodied in law and those embodied in ethical custom*” [6] Levitt in his paper “The dangers of corporate social responsibility” published in 1958 has a more critical approach compared to Friedman. According to Levitt, CSR is a “fashion accessory” in the hands of managers who don’t prioritize the wellbeing of the company they are supposed to run but instead focus on individual

goals and interests. That paper doesn't exclude the argument that CSR is and can be profitable for companies, but CSR is described as vail hiding the real purpose of these strategies: profit. If there was no profit in the background, companies wouldn't pursue any of these initiatives. [7] As we move further from the 70' when CSR was considered a vague concept, non-clearly defined and subjective to serve as a constructive part of business management strategies, studies are showing that actually CSR plays an integral part in understanding and developing new strategies in regards to how corporates work nowadays. [8] When analyzing the new characteristics of firms that have emerged in the market in the last decade, Zingales suggests that a firm should have well-defined objectives, goals, vision, mission and responsibilities. This way each company will have a clear definition of their role in the society. Changes in any of these parameters should be reflected in the business strategies of the firm in real time in order to adapt to any impact this may cause in society. Comparing in detail the traditional firms versus more contemporary firms, Zingales recommends a new theory of managing corporates given that in contemporary companies the most valuable asset is the human capital and the strategies are in a continuous change. Traditional companies had a more robust managing strategy and were oriented towards full control over assets and paid attention to hierarchy. [9] Studies suggest that consumers value a product that was produced responsibly and are willing to pay a higher price. [10] 92% of the world's 250 largest companies by revenue as defined by the Fortune 500 have generated a CSR report in 2015, confirming that CSR is established as business practice among these companies. [11] 76% of 1409 CEOs interviewed worldwide from PwC answered 'Yes' to the question 'Do you agree that business success in the 21st century will be redefined by more than financial profit?' [12] The ability of high-ranking managers to consider non-financial initiatives as part of their success is a clear indicator that sustainability and social impacts studies have outgrown in the recent year. Corporates are finding that applying CSR initiatives into their business standard is connected to employee motivation, satisfaction, and productivity. Anytime a question about sustainability and protection of environment is posed the most common answer is that governments need to invest more in creating schemes, strategies and policies that will force companies into following CSR frameworks. Business leaders, government or even researchers are mainly focused on the concept of Corporate Social Responsibility as an aspect of environment protection. In fact, studies have shown that by targeting employees and designing strategies that directly affect the workforce, sustainability increases.

Same assumption is supported by the study 'Testing theories of fairness—Intentions matter' where conclusions clearly indicate that the attribution of fairness intentions is important in both the domains of negatively and positively reciprocal behavior. [13] It is a tricky balance to get right the proportion between using CSR as an additional tool of their HR policies and genuine intentions of caring about social impact.

2. MATERIALS AND METHODS

A dataset with one thousand companies currently operating in The Czech Republic was identified using the website jobs.cz. The website of each company was checked individually in order to identify components that relate to CSR. Data collected was stored in a dataset with the following header:

Table 1. Dataset header

Company Name	Industry Sector	Company Size	Annual Turnover	Headcount	Have a CSR page	Initiatives	Link

Company Name contains the official name of the company as they are registered in Czech Republic. *Headcount* is used to determine the size of the company and it's collected from the company's website, jobs.cz or directly in Wikipedia. The headcount corresponds to the total number of full-time employees, part-time employees, seasonal workers and partners. The annual turnover is calculated using the yearly income during the reference year minus any rebates that need to be paid out.

Have a CSR page is a Boolean column where Yes or No is marked. If the company contains in their website any information regarding CSR, the entire text describing CSR is stored under column *Initiatives* together with the relevant link in the column *Link.*

If the company doesn't refer CSR in anyway in their website, then *Have a CSR* page is marked as 'No' and the two remaining columns are left empty. Simple quantitative analysis is used to determine how many companies out of one thousand have a dedicated sector in their website referring to CSR. For deeper analysis, the research work is divided into 3 main questions which will be answered in the results part:

Research question 1: *What are the keywords used on companies' websites operating in Czech Republic that convey information regarding corporate social responsibility?*

The methodology consists of 3 main steps:

1) Content filtration – Only information related to CSR was considered relevant and any other parts from the websites were not taken into account.
2) Content transformation – For each company, the text about CSR was transformed into space delimited words. Each word will be represented as node in the graph theory analysis.
3) Content reduction – Words like prefixes, suffixes, exclamations, conjunction and pronouns were removed in order to reduce connection between word classes that have no significance in CSR.
4) High data analysis – Google's Embedded Projector will be used to analyze the dataset specifically with the UNMAP algorithm.

Research question 2: *Is there a correlation between a company having a CSR webpage and industry?*

To determine the industry sector of the companies included in the dataset, Global Industry Classification Standard (GICS) was used. According to GICS, 11 industry sectors represent the majority of industry types nowadays. The classification is further detailed into 24 industry groups, 69 industry types and 158 sub-industries. [14] Since this research will use the industry sector for labelling purposes only, we will refer to the 11 industry sectors without getting deeper into the hierarchy. In order to apply quantitative analysis on the dataset, the companies included in this study were divided into these main industry sectors: [14]

- Communication Services is an industry that includes media and entertainment or any of the telecommunication services.
- Consumer Discretionary involves retails industry, hotel, restaurants, leisure and household durables.
- Consumer Staples is an industry category that groups all food products, beverages and tobacco. Energy includes oil, gas, consumable fuels and energy services.
- Financials is a category grouping all banking services, capital markets and insurance services.
- Health Care involves health care providers and pharmaceuticals.
- Industrials includes transportation services such as airlines, marine, road &rail and all services related to it.
- Information Technology involves IT services, software, technology hardware, storage and peripherals.
- Materials includes all industry sectors that produce chemicals, construction materials, packaging, metals, and mining.
- Real estate includes the real estate investment trusts and real estate services.
- Utilities includes electric, gas and water utilities services.

Research question 3: *Is the size of the company correlated with CSR?*

When creating the dataset for this research paper all companies where considered without a size restriction in order to create a full overview on how Czech companies conceptualize CSR.

To analyze the third research question, each company is assigned with an enterprise category and quantitative analysis was applied. A detailed explanation about how companies were assigned to an enterprise category will follow in sector 3.3.

2.1. Embedded Projector and UNMAP algorithm

An embedding is essentially a low-dimensional space into which a high dimensional vector can be translated. During translation, an embedding preserves the semantic relationship of the inputs by placing similar inputs close together in the embedding space. [15]

Embeddings are very useful in reducing the multidimensional spaces by helping to analyze and visualize data in clusters. In this case a cluster represents similar embeddings and the ones less connected are placed further from each other. This is archived by marking each word with their relevant distances and adding the words to the vocabulary. In this article an embedding represents the semantic relationship between words found in the websites of the companies analyzed and their connection with CSR.

Embedding Projector is an open source tool which uses the TensorFlow platform to visualize data. This platform offers 4 projections algorithm well-known for dimensionality data reduction: PCA, t-SNE, UNMAP and custom.

PCA stands for 'Principal component analysis' and it is massively used for dimensionality reduction of large datasets. It is one of the algorithms that aside of increasing the data visualization it also minimizes information loss. PCA reduces the eigenvector problem because variables are determined in hand and not upfront like some of the embedding algorithms do making it an adaptive method. [16] t-SNE stands for 'Stochastic Neighbor Embedding' and same as PCA is used for dimensionality reduction transforming large datasets into 2 or 3 dimensions. Differently from PCA that can be used for visualization and data feature extractions, t-SNE is used for data visualization only. This method calculated the distance between two embeddings after dimensionality reduction (in a low-dimension space) which makes outliers have no impact in t-SNE (differently from PCA). [17] UNMAP is an algorithm used for dimension reduction constructed from a theoretical framework based in Riemannian geometry. [18] Since the method will be used for the dataset interpretation, this research work will focus on explaining how the UNMAP algorithm is useful on interpreting the data without focusing on the theorical and mathematical interpretation. However, it is important to understand the algorithm background in order to interpret the visual and statistical results. The basis of the UNMAP algorithm is topological data analysis and simple analysis. Topological data analysis is a subject that uses mathematical functions to measure the "shape" of data and discover the interaction between cycles and clusters. These clusters can be represented by a topological network.

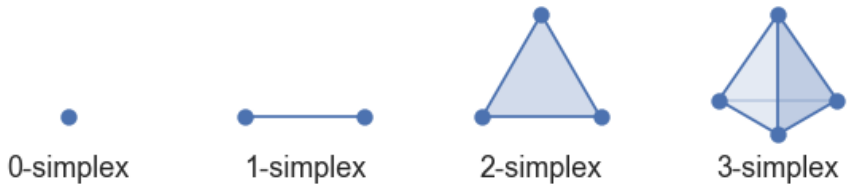
We can think of UNMAP represented by k-neighbors each having their corresponding weight. Let $X = \{x_1, \ldots, x_N\}$ be our input dataset represented as $d : X \times X \rightarrow R \geq 0$ where d is the dimensionality. For each parameter x_i we can calculate the nearest neighbor k using the nearest neighbor search algorithm integrated with the UNMAP method in Embedded Projector. For each parameter x_i, let:

$$\rho_i = \min\{d(x_i, x_{ij}) \mid 1 \leq j \leq k, d(x_i, x_{ij}) > 0\} \quad (1)$$

$$\sum_{j=1}^{k} \exp\left(\frac{-\max(0, d(x_i, x_{ij}) - \rho_i}{\sigma_i}\right) = \log_2(k) \quad (2)$$

A simplex is a straightforward method to develop a k-dimensional vertex. Hence, a 0-simplex is a point, a 1-simplex is a line two 0-simplexes, a 2-simplex is a triangle with two lines from each simplex and a 3-simplex is a tetrahedron. [19]

Figure 1. Simple dimensional simplices



Adapting this information to our real data, a cluster will represent a category of CSR. Vertices will represent one word used in the description of CSR provided by that company and edges represent the likelihood that two words are connected in the high-dimensional set. At this point, we can establish $\overline{G}$ as a weighted graph $\overline{G} = (V, E, \omega)$ where:

V represents the vertices defined in the set X, E represents the edges E= $\{(x_i, x_{ij}) \mid 1 \leq j \leq k, 1 \leq i \leq N\}$ and ω is the weight function.

$$\omega((x_i,x_{ij}))=\exp\left(\frac{-\max(0,d\left(x_i,x_{ij}\right)-\rho_i}{\sigma_i}\right) \quad (3)$$

In Embedding Projector, UNMAP provides two ways of measuring the distance between two points in space: cosine distance and Euclidean distance. Cosine distance is a measuring method that quantifies the similarity between two points in space. By using the Euclidean dot product, mathematically is expressed as:

$$\text{similarity}=\cos\theta = \frac{A\cdot B}{||A||||B||}=\frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A^2{}_i}\sqrt{\sum_{i=1}^{n} B^2{}_i}} \quad (4)$$

The resulting values indicate: [20]

Table 2. Cosine similarity value

Cosine similarity value	Indication
-1	opposite values
0	no correlation
0<x<1	intermediate similarity
1	exactly the same

2.2. Data transformation and reduction

Following the steps of the methodology explained, the content filtration of the dataset is described above with one example: **Company 1:** "*Through our commitment to social and economic progress, and our passion for bringing health through food to as many people as possible, we aim to generate profitable, sustainable growth now and for many years to come. Our mission is bringing health through food to as many people as possible. Across the world, our employees, whatever their role, are passionate and committed to leading the change to encourage more sustainable and healthier eating and drinking practices.*" [21] Content filtration and reduction after applying the simple code in TensorFlow would look like: *commitment social economic progress passion health through food people possible aim generate profitable sustainable growth years come mission health food people possible. across world employee whatever role passionate committed leading change encourage sustainable healthier eating drinking practices*

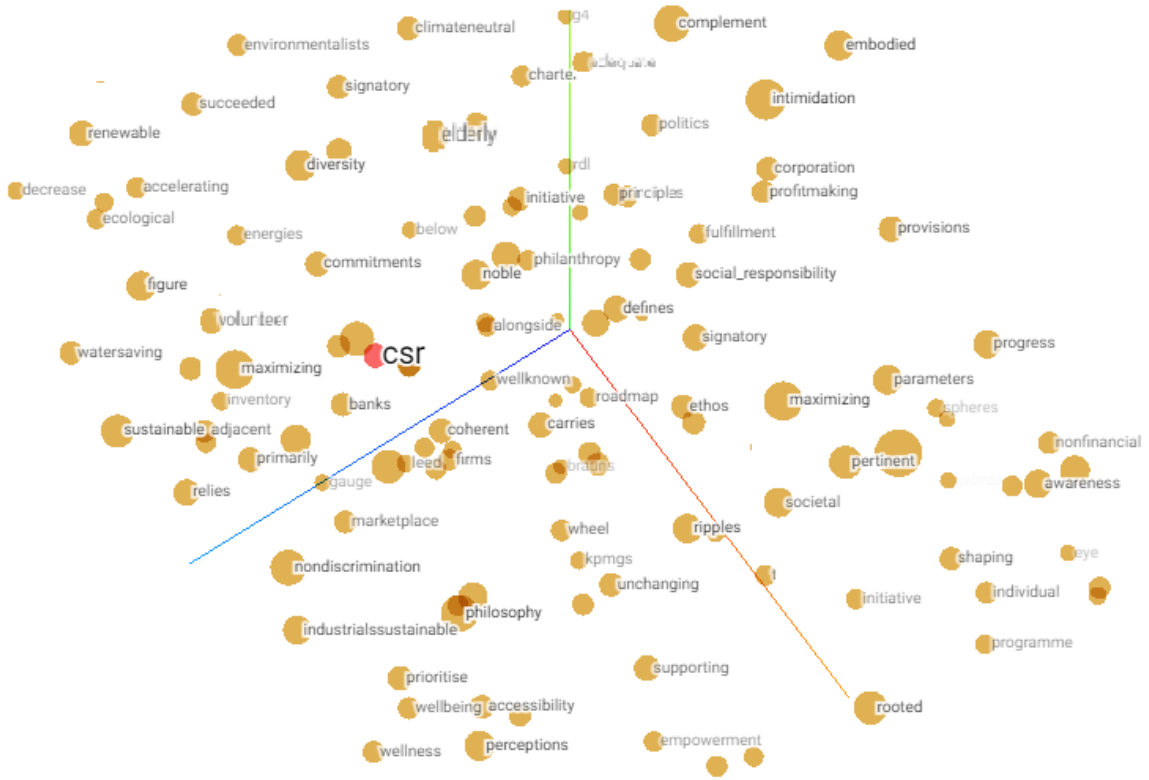
3. RESULTS

3.1. Results and discussion of research question 1

Research question 1: *What are the keywords used on companies' websites operating in Czech Republic that convey information regarding corporate social responsibility?*

After gathering, transforming, and filtering the data, the dataset was loaded in Tensorboard to be further analyzed via the UNMAP method. Given that our CSR descriptive text per each company is now space delimited to preserve the linguistic meaning, word2vec methodology will be used to create the word embeddings. Word2Vec is an algorithm that accepts text as an input and produces a coded vector for each word making natural language computer readable. [22] Being an algorithm integrated with TensorBoard, word2vec is easily accessible in python and it is a straightforward way to create word embeddings. To identify the most common used words that describe corporate social responsibility, 'CSR' abbreviation was used as a label. Data was normalized using the UNMAP algorithm and cosine distance determines the nearest points (words) in the original space. The isolated space of 100 neighbors labeled by 'CSR' is represented below:

Figure 2. CSR as a label and isolated space of 100 neighbors.



In an isolated 3D space each of the dimensions assists in creating clusters which are used to define CSR in Czech Republic.

Table 3. CSR clusters in Czech Republic

Cluster 1		Cluster 2		Cluster 3	
Word	**Cosine Distance**	**Word**	**Cosine Distance**	**Word**	**Cosine Distance**
sustainable	0.882	diversity	0.682	charity	0.645
society	0.879	volunteer	0.680	commitment	0.641
awareness	0.708	wellbeing	0.673	perceptions	0.638
energy	0.706	accessibility	0.673	roadmap	0.638
water saving	0.704	empowerment	0.67	progress	0.638
water	0.704	social	0.669	corporate	0.638
green	0.701	responsibility	0.667	political	0.634
renewable	0.700	embodies	0.666	principle	0.634
accelerating	0.699	education	0.662	initiative	0.634
ecological	0.699	donor	0.661	well-known	0.634
decrease	0.697	contributor	0.66	maximizing	0.632
environmentalists	0.695	compensation	0.658	coherent	0.632
solar	0.693	benefit	0.652	fulfillment	0.627
environment	0.692	training	0.649	innovation	0.627
plastic	0.692	aid	0.648	technology	0.616
zero waste	0.691	reward	0.648	nonfinancial	0.616
clean	0.691	compliance	0.646	individual	0.613
nature	0.690	nondiscrimination	0.646	programme	0.596
eco	0.690	intimidation	0.645		
climate	0.686	provision	0.642		
biodiversity	0.685				
neutral	0.684				

The differing industries, size of company and motivations for corporations to undertake initiatives in the three respective CSR clusters, and the wide variations in how the initiatives are managed in each cluster, underscore the challenge for corporations in crafting a unified CSR strategy. For some companies, it is entirely logical to look for measurable results and effects of CSR in their business, but for others, simply describing a broad theme around community contribution is sufficient, even if they are not quantifiable.

3.2. Results and discussion of research question 2

Research question 2: *Is there a correlation between a company having a CSR webpage and industry?*

From one thousand companies analyzed in this research work only 399 companies have a header/website/reference about CSR in their website. Each company was assigned an industry sector according to GICS methodology and the summarized data is displayed below:

Table 4. Correlation between company industry sector and CSR websites

Industry Sector	Nr. of companies	Companies with a CSR reference	%
Communication Services	48	17	35.42
Consumer Discretionary	225	91	40.44
Consumer Staples	102	33	32.35
Energy	32	15	46.88
Financials	128	28	21.88
Health Care	57	16	28.07
Industrials	130	61	46.92
Information Technology	175	110	62.86
Materials	82	25	30.49
Real Estate	18	3	16.67
Utilities	3	0	0.00
Total	1000	399	

The industry that refers CSR the most in the dataset is Information Technology with 62.86% of the companies having a website dedicated to corporate social responsibility. This is an expected result as 59% of Czech labor is directly or indirectly hired from this sector. [23]

Some of the main reasons why this industry wins over the other industry sectors are:

- The competition in this industry is high, so management needs to prepare effective communication strategies and follow CSR frameworks.
- The main asset of this industry are the employees, and this gets reflected also on the terminology used about CSR where well-being and diversity are employee-oriented only.
- Technology and Telecommunications companies are newly founded compared with the other sectors concluding to better strategies in this sector. Certainly, the management has understood the benefits of CSR and applied it correctly into their business. The industry with the least reference to CSR is Financials and Real Estate. This sector provides services and abstract products so many companies lack to refer to CSR by using the term “Sustainability” or any other terminology linked with environment. Being a well-established industry, the competition for labor is in a very stable condition and as a result there is little interest in attracting employees by referring to CSR with terminology such as well-being, diversity or community. Results and discussion of research question 3

Research question 3: *Is the size of the company correlated with CSR?*

According to Dang [24] no previous study provides a single measuring method to determine the company size. The choice of the method depends directly on the purpose of study and each measuring method will manifest advantages and disadvantages. Given this study needs the company size only for labelling purposes, 3 elements were taken into consideration when determining the company size: headcount, annual turnover, and balance sheet total. The reference year for all data gathering was 2019 as annual turnover data are sometimes published with a delay depending on the fiscal year closure. The headcount of the company was sorted in the following groups: less than 10 employees, 10-50 employees, 50-100 employees, 100-200 employees, 200-500 employees, 500-1000 employees, 1000-2000 employees, 2000-4000 employees, 4000-8000 employees and over 8000 employees. The headcount corresponds to the total number of full-time employees, part-time employees, seasonal workers and

partners. The annual turnover is calculated using the yearly income during the reference year minus any rebates that need to be paid out. The balance sheet total refers to the total value of assets of the company. [25]

Table 5. Company size category

Company Size Category	Headcount	Annual Turnover	Balance sheet
Macro	<10	≤ 2mil €	≤ 2mil €
Small	<50	≤ 10mil €	≤ 10mil €
Medium	<250	≤ 50mil €	≤ 50mil €
Large	>251	>50mil €	>50mil €

Based on this method, the companies in the dataset were divided in 4 categories: macro, small, medium, and large. For each company category, only companies that contain a CSR webpage or any reference of CSR in their website were filtered out. Results are summarized below:

Table 6. Company category and correlation with CSR

Company Size	Nr of companies in dataset	Companies with a CSR website	%
Macro	128	10	7.81
Small	210	41	19.52
Medium	334	128	38.32
Large	328	220	67.07
Total	**1000**	**399**	

As stated by Hardwick [26], there is a correlation between company size and charity/ volunteering actions. The bigger the company, the higher the percentage of the companies referring CSR in their websites. Companies that have been longer in the market, have an established market and depend on the number of employees as an import asset of their business have understood the importance and benefits of CSR in their business. This is also reflected in the way of how clearly, they communicate their initiatives into public. Benefits such as image win, mitigation of risks or cost saving are just some of the positive aspects that only companies with a clear vision and mission can incorporate into their strategies. [27] It is important to highlight that this research question does not try to link the size of a company with profitability. A big company interested in corporate social responsibility does not make more profit than a company of the same size not interested in CSR. These findings have been supported since 1985 when correlating CSR as a concern for society with profitability provided statistically non-significant results. [28] Although recent studies support the opposite, this will be further discussed in other works and does not fall in the scope of this paper.

4. DISCUSSION

The foundations of this paper are the literature review of high-quality researches and the empirical methodologies used to answer the research questions. CSR definition and frameworks are still developing, and this paper contributes to this field of study by providing an empirical measurement to the correlation of CSR with company size and CSR with industry sector, after defining the most used terminology of describing CSR in Czech Republic. The first research question finds three main clusters

that define CSR in the Czech Republic whilst emphasizing the absence of a measuring framework. Works like [29] have divided CSR in classes based on the motivation such as: economic responsibilities, legal responsibilities, ethical responsibilities, and discretionary responsibilities. Three out of four of these classes are defined as clusters in our algorithm. The reason behind this mismatch can be the dimensionality of the chosen space where UNMAP tries to squeeze everything into the 3-dimensional space. Sustainability & environment, ethics & management, and people (employees) are identified as clusters with the corresponding most used terminology to describe them. The second research question focused on finding a correlation between CSR and company industry. Studies like [30] have observed that CSR has significant value only for sectors such as health care and consumer discretionary. In the Czech Republic, the industry sector most interested in CSR resulted the technology sector. The limitation of this study is basing the research questions only on one country. Whilst this limitation assists on holding out any influences of diverse environments, it also enervates the generalization of conclusions. The third research question analyzes the correlation between company size and CSR. Conclusions support the same findings as Hardwick [26] There is an ongoing discussion in regards of this correlation and other researches may agree with Parket and Eilbirt that state: "*To be sure, the scope of endeavor categorized by the term social responsibility cannot be analyzed on the order of a balance sheet or profit and loss statement. There are, as yet no accounting techniques, analytical tools or statistical methods which will objectivity differentiate companies that are socially responsible from those that are not. To measure degrees of social responsibility would be an even more ethereal task.*" [31] This paper does not try to prove right or wrong to neither of the approaches. By relaying on methodologically strong mechanisms such as high dimensionality data algorithms and machine learning, it makes an attempt to correlate the company size with CSR without inflicting profitability as a parameter.

5. CONCLUSION

Despite the fact that information is broadly available for many organizations at once – many companies regularly fail to present the CSR data in a consistent way and assorted according to a framework. As the attention towards CSR is raising and the community becoming more watchful, the need for a standardized definition and CSR framework has arisen. This lack of standard leads to a rather flexible, unpredictable, ad hoc approach on how to deal with CSR. The purpose of this research work was to gain visibility of the current understanding of CSR in companies that operate in the Czech Republic. The findings of this work can be used to develop further frameworks and management strategies in order to better communicate CSR initiatives to stakeholders being those external or internal. By identifying the most common terminology used nowadays to refer to CSR in the Czech market, companies can focus on what customers and employees are being exposed to. The findings and implications of this analysis can inform both research and policymaking in the country. As well, this paper can be used as a guideline from managers that decide how to communicate CSR to stakeholders either internal or external.

REFERENCES

[1] Zamangi, 2006. Corporate Social Responsibility and 'Democratic Stakeholding'.

[2] Smith, J., 2012. *The companies with the Best CSR Reputations,* s.l.: Forbes.

[3] Microsoft, 2019. *Microsoft's commitment to sustainability,* s.l.: Microsoft.

[4] SAP, 2020. *Sustainability and Corporate Social Responsibility (CSR).* [Online] Available at: https://www.sap.com/corporate/en/company/sustainability-csr/csr.html [Accessed 1 06 2020]

[5] Upfield, 2020. *Sustainability and Corporate Social Responsibility*. [Online] Available at: https://upfield.com/purpose/sustainability/ [Accessed 1 06 2020].

[6] Friedman, M., 1970. The Social Responsibility of Business is to Increase its Profits. pp. 1-6.

[7] Leavitt, T., 1958. The Dangers of Social Responsibility. *Harvard Business Review,* Volume 36, pp. 41-50.

[8] Kristoffersen et al, 2005. Corporate social performance and financial characteristics: Australian evidence. *Joondalup,* 5(15), pp. 71-92.

[9] Zingales, L., 2000. In Search of New Foundations. *The journal of finance,* 55(4), pp. 1623-1653.

[10] Bhattacharya et al., 2010. Corporate Social Responsibility and Competitive Advantage: Overcoming the. *Management Science,* 57(9), pp. 1528-1545.

[11] KPMG, 2016. *The KPMG Survey of Corporate Responsibility Reporting 2016,* s.l.: KPMG.

[12] PwC, 2016. *Redefining business success in a changing world: CEO survey,* s.l.: PwC.

[13] Falk et al, 2008. Testing theories of fairness - Intentions matter. *Games and Economic Behavior,* 62(1), pp. 287-303.

[14] MSCI, 2020. *Global industry classification standard (GICS) methodology,* s.l.: MSCI.

[15] Pandey, P., 2019. *Towards Data Science.* [Online] Available at: https://towardsdatascience.com/visualizing-bias-in-data-using-embedding-projector-649bc65e7487 [Accessed 15 12 2020].

[16] Ian T. Jolliffe, J. C., 2016. Principal component analysis: a review and recent developments. *Royal Society,* 13 4.374(2065).

[17] Laurens van der Maaten, G. H., 2008. Visualizing Data using t-SNE. *Journal of Machine Learning,* p. 27.

[18] McInnes et al., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv.org,* p. 51.

[19] McInnes, L., 2018. *https://umap-learn.readthedocs.io/.* [Online] Available at: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html [Accessed 06 06 2020].

[20] NIST, 2017. *National Institute of Standards and Technology.* [Online] Available at: https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm [Accessed 18 12 2020].

[21] Danone, 2020. *Danone.com.* [Online] Available at: https://www.danone.com/about-danone/sustainable-value-creation.html [Accessed 15 12 2020].

[22] Mikolov, T., 2013. Efficient Estimation of Word Representations in Vector Space. *Cornell University,* Volume v1.

[23] Večerník, J. V. F., 2015. *The Labour Market in the Czech Republic: Trends, Policies and Attitudes,* s.l.: s.n.

[24] Dang, C. &. L. F. &. Y. C., 2017. Measuring Firm Size in Empirical Corporate Finance. *Journal of Banking & Finance,* Issue 10.1016/j.jbankfin.2017.09.006, p. 86.

[25] ECHA, 2020. *How to determine the company size category,* s.l.: ECHA.

[26] Hardwick et al., 1998. An Analysis of Corporate Donations: United Kingdom Evidence. 35(5), pp. 641-654.

[27] Schieg, M., 2009. *The Model of Corporate Social Responsibility in Project Management,* s.l.: Researchgate.com.

[28] Hatfield et al., 1985. An Empirical Examination of the Relationship between Corporate Social Responsibility and Profitability. *The Academy of Management Journal,* 28(2), pp. 446-463.

[29] Carroll, A. B., 1979. A Three-Dimensional Conceptual Model of Corporate Performance. *The Academy of Management Review,* 4(4), pp. 497-505.

[30] Hoepner et al., 2010. Corporate Social Responsibility Across Industries: When Can Who Do Well by Doing Good?. *SSRN*, 57(3), pp. 18-41.

[31] Eilbirt et al., 1973. The practice of business: The current status of corporate social responsibility. *Business Horizons*, 16(5), pp. 5-14.