CSRCZ：捷克共和国企业社会责任数据集

Xhesilda Vogli
捷克生命科学大学经济与管理学院管理系 vogli@pef.czu.cz

Erion Çano
数字文献学、数据挖掘与机器学习，奥地利维也纳大学
erion.cano@univie.ac.at

arXiv:2301.03404v1 [econ.GN] 5 Jan 2023

摘要

随着利益相关者对企业披露其企业社会责任实践的压力日益增大，了解企业信息披露系统在弥合企业社会责任报告与实际实践之间差距方面的效率至关重要。与此同时，企业社会责任研究仍未跟上近期数据驱动策略的发展步伐，可获取的公开数据也很少。本文旨在介绍 CSRCZ，这是一个新近创建的数据集，基于在捷克共和国运营的 1000 家公司的官方网站所发布的披露报告构建。我们依据三个主要参数分析了每家公司：公司规模、公司行业和公司倡议。本文介绍了该数据集的内容及其在未来研究中的潜在用途。我们认为，CSRCZ 对后续研究具有重要意义，因为它是首个公开发布的同类数据集。

1 引言

企业社会责任（CSR）从 20 世纪 50 年代初的一个“为什么”（Carroll and Brown），2018年）发展至近年来的“必需”。一般而言，CSR 被视为一种自我规制型商业模式，有助于公司为社会目标作出贡献，并对自身及公众承担社会责任。CSR 受到公司运营所在国家的法律环境（LIANGand RENNEBOOG，2017）和社会政治环境（Tilt，2016）的显著影响。在全球范围内，越来越多的公司开始参与企业社会责任倡议，因此向公众提供了更多社会信息。因此，企业社会责任信息披露已成为该领域研究者的主要研究方向之一（Goyal 等，2015；Halkosa and Skouloudis，2016）。

尽管达到可持续发展信息披露或报告的适当标准值得提倡，但仍有若干障碍需要克服。与受到严格监管的财务报告不同，可持续发展报告是自愿的，因此缺乏统一性（Braamand Peeters，2018年；Bhattacharyya and Cummings，2015年）。以往研究通常聚焦于推动此类举措披露的因素、所提供的信息、沟通方式，以及这些举措对公司绩效和形象的影响（Gonçalvesand Gaio，2023年；Benoit-MoreauandParguel，2011年）。可能影响公司企业社会责任信息披露报告的因素通常分为：（i）内部因素，例如公司规模、行业领域、财务绩效和公司治理；（ii）外部因素，例如原属国、利益相关者、媒体，或社会与政治环境（Fifka，2013年；Morhardt，2009年）。

鉴于现有研究有限，也有少数研究尝试探讨“国家”是否会影响企业社会责任倡议和披露水平（Kansal 等人，2014年；Fufa 和 Roba，2021年；Khan 等人，2021年）。一方面，其他因素与公司企业社会责任倡议之间更深层次的相关性大多尚未得到研究。另一方面，大多数研究（例如上述研究）在方法上较为“保守”，没有采用过去十年兴起的数据驱动方法（Pugna 等人，2022年；Abuimara 等人，2022年；Çano，2018年）。这种数据驱动研究趋势主要借助英语资源（例如数据集）开展，因为这类资源在互联网上数量最多。不过，捷克语或其他语言的研究和资源仍在不断涌现并日益普及（Çano 和 Bojar，2019年；Sestino 和 Mauro，2022年）。

本文旨在通过创建并介绍 CSRCZ，推动基于数据的企业社会责任研究。CSRCZ 是一个免费开放的数据集，包含在捷克共和国运营的 1000 家公司的公开信息。[1] 在接下来的章节中，我们将介绍所采用的信息检索流程步骤，并描述可用的

[1] https://zenodo.org/record/7495802

属性	内容类型
公司名称	字符串
员工人数	整数
是否有 CSR 页面	二元
行业领域	字符串
公司规模	分类变量
举措	字符串
网站	URL

表 1：数据属性及其对应的类型。

数据字段（尤其是与企业社会责任相关的字段）、其特征值以及一些相关统计数据。最后，我们将讨论在未来企业社会责任研究情境下，CSRCZ 内容的潜在用途。

2 数据集内容

构建 CSRCZ 数据集所用的数据来源于目前在捷克共和国运营的 1,000 家公司的公共网站。首先，通过 jobs.cz 检索这些公司的网站。随后对每个网站进行分析，仅收集与企业社会责任相关的信息。表 1 展示了所考察的相关属性。

公司名称 表示公司在捷克共和国注册的正式名称，以文本字符串形式保存。员工人数 是一个整数，包含全职员工、兼职员工、季节性工人和合作伙伴的总人数。是否有 *CSR* 页面 是一个二元值，其中“1”表示该公司网站上有介绍企业社会责任政策或实践的页面，“0”则表示没有。行业领域 是一个字符串，用于描述公司的市场细分或其主要从事的活动类型。

公司规模 是一个分类变量，用于描述公司的规模。员工人数少于 10 人的公司被视为“微型”。员工人数不超过 50 人的公司属于“小型”公司。员工人数在 51 至 250 人之间的公司被视为“中型”。员工人数达到或超过 251 人的公司属于“大型”。举措 可能是企业社会责任分析中最重要的属性。它是一个较长的字符串，用于描述公司列出的任何与企业社会责任相关的政策、实践或举措。最后，网站 是检索信息所用的统一资源定位符。

Size	数量	百分比
未知	1	0.1
微型	125	12.5
小型	214	21.4
中型	330	33
大型	330	33

表2：所选公司的规模统计

3 数据集统计

以下各节将详细讨论 CSRCZ 的内容。我们将分析相关字段的特征值，并以表格形式呈现。用于生成统计数据的代码可在线获取。[2]

3.1 规模和员工

公司的规模是一个重要因素，通常与公司实施目标和实践以履行其企业社会责任战略的能力有关。确定公司规模的一种方法是使用员工人数，这与我们在第2节中描述的方法相同。这显然是一种简单化的方法，因为公司拥有的不同类型资产等其他因素（遗憾的是，这类信息并不总是公开的）也能表明公司的规模。

我们检查了收集到的数据，发现大多数公司属于大型或中型，两类各占样本的 33 %。此外还有 214 家小型公司，占总数的 21.4 %。另有 125 家公司（占总数的 12.5 %）被归类为微型。对于其中一家抽样公司，无法确定其规模。完整统计数据见表2，并见图1。

我们还统计了各规模类别的员工人数。具体而言，我们计算了 CSRCZ 数据中“微型”“小型”“中型”和“大型”公司的员工人数最小值、最大值和平均值。“微型”公司至少有 5 名员工，最多有 9 名，平均为 6.54 名。“小型”公司的相应统计值分别为 10、49 和 31.97。“中型”公司平均有 169.62 名员工。最后，“大型”公司的员工人数最大值确实达到 10000 人（数据中最大的

[2] https://github.com/erionc/csrcz-stats

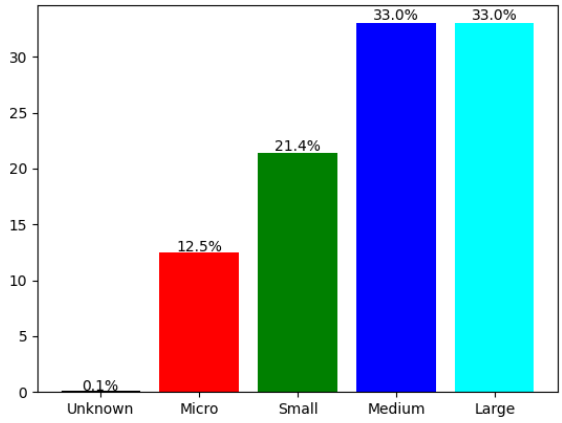


图1：所选公司的规模分布。

公司	Min	Max	Avg
微型	5	9	6.54
小型	10	49	31.97
中型	50	249	169.62
大型	299	10000	1635.58

表3：各公司类别的员工人数最小值、最大值和平均值。

CSRCZ）的平均值为 1635.58。统计数据汇总于表3，并显示在图3中。

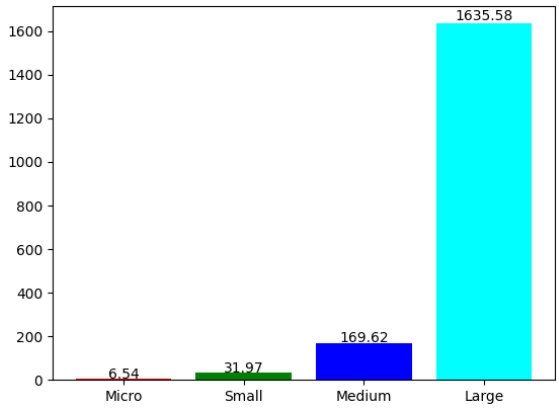


图2：各公司规模类别的员工平均人数。

3.2 行业领域

行业领域是一项有趣的属性，因为它有助于揭示与企业社会责任倡议以及公司所处不同行业相关的重要趋势。根据全球行业分类标准（GICS），如今大多数行业类型可归为十一个行业领域。[3]

[3]https://www.msci.com/our-solutions/ indexes/gics

行业领域	数量	百分比
未知	607	60.7
通信服务	17	1.7
非必需消费品	91	9.1
必需消费品	31	3.1
能源	15	1.5
金融	28	2.8
医疗保健	16	1.6
工业	111	11.1
信息技术	56	5.6
材料	25	2.5
房地产	3	0.3
公用事业	0	0

表4：所选公司的行业统计数据

通信服务 是一个涵盖媒体和娱乐或各类电信服务的行业。

非必需消费品 涉及零售业、酒店、餐饮业、休闲和耐用品。

必需消费品 是一个涵盖所有食品、饮料和烟草的行业类别。

能源 包括石油、天然气、消耗性燃料和能源服务。

金融 是涵盖所有银行服务、资本市场和保险服务的类别。

医疗保健 涉及医疗保健服务提供商和制药业。

工业 包括航空公司、海运、公路和铁路等运输服务，以及与之相关的所有服务。

信息技术 涉及信息技术服务、软件、技术硬件、存储设备和外围设备。

材料 包括所有生产化学品、建筑材料、包装材料、金属和采矿产品的行业部门。

房地产包括房地产投资信托和房地产服务。

公用事业包括电力、天然气和水务公用事业服务。

企业社会责任倡议	Min	Max	Avg
字符	0	32023	1218.01
词元	0	4870	191.73

表 5：每项企业社会责任倡议的字符数和词元数的最小值、最大值及平均值。

我们对数据进行了探索，并统计了属于上述各行业部门的公司数量及其占比。收集到的统计数据汇总于表4中。遗憾的是，许多数据记录都没有这一指标。在可获取的行业部门数据中，“工业”最为常见，共有 111 家公司，占总数的 11.1%。其次是“非必需消费品”，共有 91 家公司。“ 信息技术 ”、“ 必需消费品 ”和“金融”也较为常见，分别有 56、31 和 28 条记录。最少见的行业部门是“房地产”和“公用事业”，分别有 3 家和 0 家公司。

3.3 企业社会责任倡议

CSRCZ数据集中最重要的记录属性或许是“举措”，其中概述了各公司的企业社会责任使命、目标和实践。这些信息通常以一连串句子的形式呈现，有时也会由几个段落组成。一个简单的统计评估方法是检查其字符数或词元数，尽管网站上的“举措”文本较短或较长，并不一定意味着公司的企业社会责任承诺较低或较高。

我们使用 NLTK词语分词器对文本进行分词。[4] 遗憾的是，抽样公司中有相当多的公司（具体而言，有 610 家公司的字符数和词元数均为 0）未在其网站上提供此类描述。最长的企业社会责任倡议文本包含 32023 个字符和 4870 个词元。此属性的平均长度约为 1218 个字符和 191 个词元。这些统计数据汇总于表5。

4 讨论

尽管许多组织都能广泛获取相关信息，但许多公司仍经常无法以一致的方式呈现企业社会责任数据，也未能按照框架对其进行分类。随着人们对企业社会责任的关注度不断提高，公众也愈发审慎，因此，对标准化定义和企业社会责任框架的需求日益增长。应用数据驱动方法并提供结构化数据集的需求也在不断增加。

这项工作的目的是通过提供并介绍近期创建的数据集 CSRCZ，促进以数据为驱动的企业社会责任研究。我们认为，使用 CSRCZ 有助于更好地了解在捷克共和国运营的公司对企业社会责任的当前认知，也能从全球情境的角度加深理解。我们可以发现公司内部和外部因素与其企业社会责任倡议之间的各种相关性。这些发现可用于进一步制定框架和管理策略，以便更好地向外部或内部利益相关者传达企业社会责任倡议。

[4] https://www.nltk.org/

参考文献

Tareq Abuimara、Brodie W Hobson、Burak Gunay 和 William O’Brien。2022年。提升商业建筑节能运行的数据驱动工作流程：结合真实案例的综述。《建筑服务工程研究与技术》，43(4)：517–534。

Florence Benoit-Moreau 和 Béatrice Parguel。2011年。通过环境传播构建品牌资产：法国的一项实证研究。《欧罗美德商业期刊》，6：100–116。

Asit Bhattacharyya 和 Lorne Cummings。2015年。衡量企业环境绩效——利益相关方参与评估。《商业战略与环境》，24(2)：309–325。

Geert Braam 和 Roy Peeters。2018年。企业可持续发展绩效与可持续发展报告鉴证：可持续发展领域会计实践的传播。《企业社会责任与环境管理》，25(2)：164–181。

Erion Çano。2018年。基于文本的情感分析与音乐情感识别。博士论文，计算机工程，都灵理工大学，意大利都灵。

Erion Çano 和 Ondˇrej Bojar。2019年。捷克语文本的情感分析：一项算法综述。载于第*11*届国际智能体与人工智能会议论文集——第*2*卷：*NLPinAI*，第973–979页。INSTICC，SciTePress。

A.B. Carroll 和 J.A. Brown。2018年。企业社会责任：当前概念、研究与议题综述。《国际企业社会责任期刊》（*JCSR*），2(2):39–69。

Matthias S. Fifka。2013年。《企业责任报告及其决定因素的比较视角——实证文献综述与元

分析。《商业战略与环境》，22:1–35。

Tolossa Fufa 和 Yadessa Roba。2021年。发展中国家跨国企业子公司企业社会责任实践的内部和外部决定因素：来自埃塞俄比亚的证据。《未来商业期刊》，7。

Tiago Cruz Gonçalves 和 Cristina Gaio。2023年。企业可持续发展信息披露与媒体曝光度：来自旅游业的混合方法证据。《商业研究期刊》，155:113447。

Praveen Goyal、Zillur Rahman 和 Absar Ahmad Kazmi。2015年。运用层次分析法识别并确定企业可持续发展实践的优先级。《管理建模期刊》，10(1):23–49。

George Halkosa 和 Antonis Skouloudis。2016年。探究企业气候变化信息披露的现状及关键决定因素：来自希腊商业领域的证据。《环境科学与政策》，56(1):22–31。

Monika Kansal、Mahesh Joshi 和 Gurdip Singh Batra。2014年。企业社会责任信息披露的决定因素：来自印度的证据。《会计学进展》，30(1):217–229。

Majid Khan、James Lockhart 和 Ralph Bathurst。2021年。企业社会责任的制度分析：来自一个新兴国家的经验。《新兴市场评论》，46:100752。《新兴市场中的企业社会责任》。

Hao Liang 和 Luc Renneboog。2017年。论企业社会责任的基础。《金融学杂志》，72(2):853–910。

J. Emil Morhardt。2009年。互联网上的企业社会责任与可持续发展报告。《商业战略与环境》，19:436–452。

Irina Bogdana Pugna、Dana Maria Boldeanu、Mirela Gheorghe、Gabriel Cozgarea 和 Adrian Nicolae Cozgarea。2022年。能源行业对数据驱动型组织的管理视角。《能源》，15(16)。

Andrea Sestino 和 Andrea De Mauro。2022年。在商业中利用人工智能：影响、应用与方法。《技术分析与战略管理》，34(1):16–29。

CarolA. Tilt。2016年。企业社会责任研究：情境的重要性。《国际企业社会责任期刊》（*JCSR*），1(2):1–9。