



# Mapping Keywords and Themes of UGM's SDGs Articles: A Corpus-Based Study

Erlin Estiana Yuanti<sup>1\*</sup> Tri Nuraniwati<sup>2</sup> and Azizah Ayu Nursiami Sugiharto<sup>3</sup>

<sup>1,2,3</sup> Department of Languages, Arts and Culture Management, Vocational School, Universitas Gadjah Mada Yogyakarta, Indonesia  
erlin\_estiana@ugm.ac.id

**Abstract.** This study analyzes Universitas Gadjah Mada's (UGM) news coverage of the Sustainable Development Goals (SDGs) published on [ugm.ac.id](http://ugm.ac.id) between January 2020 and June 2024. The research maps the keywords and thematic orientations of UGM's news coverage of SDGs, with a specific focus on SDGs 7, 8, 9 and 17. Using a corpus-based approach and log-likelihood keyness analysis through *AntConc 4.2.0*, the study compares the lexical prominence of terms in UGM's SDG corpus with Elsevier's SDG keyword list as a reference corpus. The findings reveal distinct thematic emphases: SDG 7 (*Affordable and Clean Energy*) centers on *energy, solar, wind, and hydro*, emphasizing renewable energy research; SDG 8 (*Decent Work and Economic Growth*) highlights *tourism, labour, employment, and productivity*, indicating focus on inclusive economic progress; SDG 9 (*Industry, Innovation and Infrastructure*) is represented by *manufacturing, industrial, infrastructure, and innovation*, showcasing technological advancement and industrial capacity building; and SDG 17 (*Partnerships for the Goals*) features *trade, debt, investment, and agreement*, underscoring the university's engagement in global partnerships and policy collaboration. Overall, the study provides a lexical and thematic map of UGM's SDG discourse, demonstrating how institutional research communication aligns with the UN 2030 Agenda through innovation, inclusivity, and international cooperation.

**Keywords:** Sustainable Development Goals, Corpus Linguistics, Keyness Analysis, Lexical Mapping, Universitas Gadjah Mada, Academic Discourse

## 1 Introduction

The 17 Sustainable Development Goals (SDGs) of the 2030 Agenda for Sustainable Development, endorsed at the UN Summit in 2015, officially took effect on January 1, 2016 [25]. The SDGs provide a comprehensive framework to address global social, economic, and environmental challenges. Higher education institutions play an important role in advancing the SDGs by contributing through their Tri dharma functions—education, research, and community service—as well as through outreach activities and institutional initiatives. Universities are increasingly positioned as key actors in advancing the United Nations Sustainable Development Goals (SDGs) through research, education, and community engagement. While the volume of SDG-related

© The Author(s) 2026

M. A. Norazlina et al. (eds.), *Proceedings of the 4th International Conference of Research on Language 2025 (IROLE 2025)*, Advances in Social Science, Education and Humanities Research 1009,  
[https://doi.org/10.2991/978-2-38476-563-8\\_26](https://doi.org/10.2991/978-2-38476-563-8_26)

academic publications continues to grow, existing studies predominantly assess institutional SDG engagement through policy analysis, bibliometric indicators, or thematic classifications. These approaches provide useful macro-level insights but offer limited understanding of how SDGs are linguistically constructed and emphasized within academic discourse.

In the higher education sector, sustainability reporting has emerged as a tool for demonstrating accountability and stakeholder engagement. However, Ceulemans et al. noted that sustainability reporting in universities remains in an evolving phase, with limited implementation and several obstacles, including inconsistent standards, insufficient stakeholder involvement, and weak integration into organizational strategies [7]. Beyond formal reporting, universities increasingly rely on news publications and online platforms to showcase their contributions. Times Higher Education (THE) Impact Rankings, for instance, consider a variety of indicators, including news publications, to assess institutional engagement with the SDGs [24].

Universitas Gadjah Mada (UGM), one of Indonesia's leading universities, has demonstrated strong commitment to sustainable development. It has established a dedicated SDGs Center and developed an online portal (*sustainabledevelopment.ugm.ac.id*) to disseminate SDG-related news, reports, and activities. These initiatives complement institutional mechanisms such as the World Class University (WCU) program, the Greenmetric Team, and the THE Impact Team. UGM's performance in the THE World University Rankings 2023 placed it second nationally after the University of Indonesia, and within the global range of 1201–1500, with notable strengths in SDGs 7, 8, 9, and 17 [18].

While sustainability reporting in higher education has been examined, little attention has been given to linguistic perspectives. Previous research on SDGs in higher education has largely relied on bibliometric and thematic approaches to examine publication trends, collaboration networks, and dominant research clusters related to sustainability and development goals. Such studies have been valuable in demonstrating the increasing engagement of universities with the SDGs agenda and in identifying institutional contributions at a macro level. Most prior studies approach SDG reporting from management or policy perspectives [9, 15]. Some corpus-based studies have been applied to corporate sustainability contexts, such as Smeuninx et al. [21] who analyzed readability in sustainability reports, and Cui [8], who investigated Apple's SDG and environmental reports using critical discourse analysis. Hou and Ayers (2023) explored SDGs in community colleges using corpus linguistics and discourse analysis [11]. However, few studies have applied corpus methods to university-based SDG news coverage. In particular, there remains a lack of corpus-based studies that systematically examine SDG-specific lexical patterns and explicitly link them to individual SDG targets, especially in the context of higher education institutions in the Global South. As a result, it is still unclear how universities articulate their SDG commitments through language use and which dimensions of the SDGs are most prominently foregrounded in institutional research outputs.

To address this gap, the present study employs a corpus-based keyness analysis to map keywords and themes in Universitas Gadjah Mada's (UGM) SDG-related publications, with a focus on SDGs 7, 8, 9, and 17. By comparing UGM's corpus with an SDG keyword reference corpus, this study aims to reveal how institutional research discourse aligns with specific SDG targets and contributes to the broader SDG agenda..

The objectives of this study are: (1) to identify the thematic focus of UGM's SDG news coverage and (2) to analyze lexical salience of Sustainable Development Goals (SDGs) 7, 8, 9, and 17 within Universitas Gadjah Mada's (UGM) SDG-related publications, benchmarked against the Elsevier SDG keyword corpus.

By adopting a corpus linguistic approach, this study contributes to both sustainability communication and discourse analysis in higher education. It provides insights into how UGM represents its SDG engagement through news publications, highlights underrepresented themes such as *decent work* in SDG 8, and offers recommendations for strengthening strategic communication. More broadly, the findings suggest a replicable methodology for assessing how universities communicate their SDG achievements, thereby supporting evidence-based strategies for advancing sustainability in higher education.

## 2 Method

This study was conducted at the Vocational School of Universitas Gadjah Mada (UGM) from April to October 2025, involving three lecturers and one student from the Bachelor of Applied English study program as the research team. Their roles included corpus compilation, keyword validation, and data analysis.

The primary data consisted of a target corpus comprised of news articles published on UGM's official website (<https://ugm.ac.id>) between January 2020 and June 2024. Articles were selected based on their relevance to the Sustainable Development Goals (SDGs), particularly SDG 7 (Affordable and Clean Energy), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 17 (Partnerships for the Goals). These SDGs were chosen because they are among the goals where UGM scored highly in the 2023 Times Higher Education (THE) Impact Rankings, as reported by the UGM World Class University (WCU) Team. A total of 81,152 tokens were compiled into a corpus by converting the articles into .txt format.

For comparison, a reference corpus was also built from Elsevier's SDG-related keywords, which are officially used in the THE Impact Rankings framework. It functions as a conceptual baseline rather than a narrative corpus. These keywords were accessed from the dataset of the Elsevier [6] which is converted into .txt format for comparability. Therefore, the keyness results indicate not only lexical overrepresentation but also the degree to which SDG concepts are discursively elaborated in institutional communication. Words with high keyness in the target corpus reflect areas where sustainability themes are actively contextualized, explained, and promoted, rather than merely acknowledged as conceptual categories. This enabled mapping and cross-referencing of UGM's reported activities with internationally recognized SDG indicators.

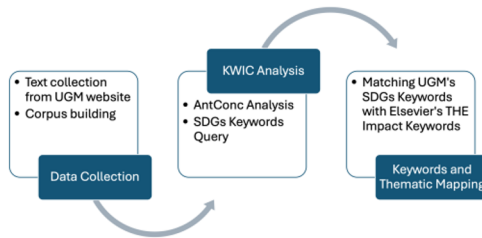
To identify the lexical items that are statistically salient in the UGM SDG-related news corpus, a keyness analysis was conducted using the log-likelihood (LL) statistical measure. The analysis compared the target corpus (UGM SDG news articles published from January 2020 to June 2024) with a reference corpus consisting of SDG-related keyword lists compiled from Elsevier's SDG Mapping Initiative. The purpose of this procedure was to determine which lexical items occur significantly more frequently in the UGM corpus than would be expected when compared with the broader conceptual vocabulary associated with each SDG objective.

The analysis was carried out using AntConc 4.2.0 [1], employing the Keyword List function with log-likelihood as the statistical test. Log-likelihood was selected because it is robust for comparing corpora of different sizes and does not assume a normal distribution of word frequency [17]. In corpus linguistics, LL is widely used to identify key words—terms that are unusually frequent in a corpus relative to another [2; 19]. In this study, a higher LL value indicates a stronger association of the keyword with the UGM corpus, suggesting that the term is more discursively foregrounded within UGM's institutional communication of SDGs.

The keyness output provided two central measures:

1. Log-Likelihood Score, indicating the statistical strength of difference; and
2. Effect Size, indicating the magnitude of relevance of that difference [12].

Only the highest-ranked keywords were selected for further qualitative interpretation, allowing the analysis to focus on lexical items that are not merely present but ideologically and thematically prominent in UGM's SDG discourse [4]. Below is the flowchart of the research:



**Fig. 1.** The flow chart of the research

### 3 Findings

The analysis identified a total of 170 UGM news articles related to four SDGs: 37 articles for SDG 7 (Affordable and Clean Energy), 14 articles for SDG 8 (Decent Work and Economic Growth), 71 articles for SDG 9 (Industry, Innovation, and Infrastructure), and 48 articles for SDG 17 (Partnership for the Goals). The full metadata of these articles is available at [ugm.id/METADATASDGS]. In the following subchapters, discussion on the finding of each SDG is presented.

#### 3.1 Findings on SDG 7: Affordable and Clean Energy

From the comparison done between the UGM SDG 7 articles and the Elsevier keywords of SDG 7, the target corpus has 122 unique keyword types compared to 1,297 unique word types in total, in the reference corpus. Also, the target corpus has 3,145 keyword tokens, compared to 5,701 total occurrences in the reference corpus. This suggests that UGM SDG articles on SDG 7 highlight 122 distinct thematic keywords and those themes occur repeatedly (3,145 times), meaning the narrative is consistent, not random.

The keyness analysis for SDG 7 compares UGM's SDG news discourse (target corpus) with the Elsevier SDG 7 keyword inventory (reference corpus). As the reference corpus consists of controlled lexical descriptors rather than narrative text, keyness values indicate not merely lexical overuse but the degree to which SDG 7 concepts are discursively elaborated and contextualized in institutional communication. High keyness therefore signals discursive prioritization, highlighting the themes that UGM actively frames and promotes as part of its sustainability identity. Below are the top twenty SDG 7-related keywords that show the highest keyness values when comparing the UGM news corpus to the Elsevier reference list.

**Table 1.** Log-likelihood values of Top 20 SDG 7 keywords (UGM vs. Reference Corpus)

No	Keyword	Frequency in UGM Corpus	Frequency in Reference Corpus	Log-likelihood
1	<i>energy</i>	313	218	267.431
2	<i>wind</i>	101	2	256.746
3	<i>power</i>	163	56	229.212
4	<i>solar</i>	122	25	213.446
5	<i>hydro</i>	70	0	190.516
6	<i>smart</i>	53	0	144.129
7	<i>tidal</i>	51	0	138.677
8	<i>turbine</i>	48	1	121.329
9	<i>hybrid</i>	59	7	119.984
10	<i>wave</i>	38	0	103.263
11	<i>bio</i>	49	5	102.866
12	<i>dam</i>	37	0	100.541
13	<i>fuel</i>	71	25	97.881
14	<i>hydroelectric</i>	36	0	97.819
15	<i>efficient</i>	35	0	95.097
16	<i>low</i>	47	6	93.89
17	<i>heat</i>	34	0	92.376
18	<i>saving</i>	39	2	91.188
19	<i>geothermal</i>	32	0	86.933
20	<i>ocean</i>	32	0	86.933

As shown in the table above, *energy* is the most statistically salient term in the corpus (LL = 267.431), confirming that SDG 7 functions as a central organizing theme within the sustainability discourse. Beyond this general lexical anchor, a cluster of terms related to renewable and alternative energy technologies also appears with high keyness scores, including *wind* (LL = 256.746), *power* (229.212), *solar* (213.446), *hydro* (190.516), *tidal* (138.677), and *turbine* (121.329). These words are present across all texts in the target corpus (Range = 1.000), indicating that renewable energy is not treated as a peripheral or occasional topic, but as a consistently narrated institutional focus. The findings above can also be linked to Target 7.1 (Affordable and Clean Energy Access). The recurrent use of the word *energy*, *power*, and *fuel* suggests ongoing institutional discourse around ensuring energy access and transitioning away from conventional fuel sources.

Importantly, several of these terms (e.g., *hydro*, *tidal*, *geothermal*, *wave*, *ocean*) appear with zero recorded instances in the reference keyword list, despite being recognized concepts within global SDG literature. Their prominence in the target corpus

demonstrates that UGM's SDG 7 discourse adopts a contextualized and implementation-oriented approach, highlighting specific technologies, local resource potentials, and ongoing projects rather than merely invoking sustainability as a general institutional value. The repeated presence of *solar*, *wind*, *hydro*, *geothermal*, *hydroelectric*, and *tidal* keywords indicate strong emphasis on renewable energy diversification and research initiatives supporting low-carbon energy infrastructures which is closely related to the Target 7.2 (Increasing the Share of Renewable Energy).

The findings also reveal a secondary lexical field centered on efficiency and systems optimization, represented by terms such as *smart*, *efficient*, *low*, *heat*, and *saving*. This suggests that the discourse does not only focus on *energy production* but also frames SDG 7 in terms of its Target 7.3 (Energy Efficiency Improvement), modernization of infrastructure, and responsible consumption behaviours. The high keyness of advanced technical terms (e.g., *turbine*, *hybrid*, *wave*) demonstrates UGM's role in positioning itself as a research leader contributing to innovation and technology development in renewable energy systems, which is relevant to Target 7.a (Research on Technology Advancement). Thus, SDG 7 is articulated as UGM's narrative on improving energy use efficiency and promoting energy conservation technologies in campus or research settings.

Taken together, the likelihood results indicate that UGM constructs SDG 7 as a site of practical, solution-oriented sustainability work. UGM discourse narrates SDG 7 through examples of local implementation, capacity development, and technological innovation. This demonstrates that the institution frames its sustainability identity around applied, community-relevant energy transformation, positioning itself as an active agent in the transition toward affordable and clean energy, rather than as a passive participant in global policy rhetoric. The consistently high log-likelihood values indicate that these terms occur significantly more often in the UGM corpus, suggesting strong thematic emphasis on sustainable and renewable energy development. Core energy terms such as *energy*, *power*, and *fuel* appear alongside more specific technological lexicon including *solar*, *wind*, *hydro*, *turbine*, *geothermal*, and *hydroelectric*, indicating that UGM's discourse frames SDG 7 largely through the lens of applied energy solutions and infrastructure. The presence of terms such as *efficient*, *smart*, and *saving* also reflects a narrative focus on optimization, energy management, and the integration of intelligent technologies. Overall, the discourse highlights UGM's alignment with solution-driven, science-based sustainable energy strategies rather than purely advocacy or policy communication.

### 3.2 Findings on SDG 8: Decent Work and Economic Growth

From the comparison done between the UGM SDG 8 articles and the Elsevier keywords of SDG 8, the target corpus has 52 unique keyword types compared to 568 unique word types in total, in the reference corpus. This means about 9.15% of all unique words in the target corpus are keywords. Also, the target corpus has 731 keyword tokens, compared to 1,767 total occurrences in the reference corpus. This suggests that about 41% of the total words in UGM SDG 8 corpus are made up of these keywords. This indicates that UGM SDG 8 corpus are strongly themed, because a small number of words (52)

occur repeatedly and make up a large portion of the entire text (731 tokens). This suggests clear topical focus and consistent vocabulary, which is typical in institutional sustainability communication.

**Table 2.** Log-likelihood values of Top 20 SDG 8 keywords (UGM vs. Reference Corpus)

No	Keyword	Frequency in UGM Corpus	Frequency in Reference Corpus	Log-likelihood
1	<i>tourism</i>	48	2	143.599
2	<i>labour</i>	38	0	126.143
3	<i>employment</i>	39	1	120.545
4	<i>child</i>	27	0	89.49
5	<i>labor</i>	26	1	78.03
6	<i>wage</i>	23	0	76.19
7	<i>trafficking</i>	21	0	69.545
8	<i>growth</i>	41	19	69.173
9	<i>productivity</i>	26	3	68.134
10	<i>trade</i>	23	2	63.093
11	<i>credit</i>	21	1	61.83
12	<i>children</i>	17	0	56.267
13	<i>technological</i>	16	0	52.95
14	<i>women</i>	18	3	43.627
15	<i>slavery</i>	13	0	43.004
16	<i>income</i>	12	0	39.69
17	<i>economic</i>	34	28	39.174
18	<i>rights</i>	13	1	36.222
19	<i>policy</i>	16	4	34.624
20	<i>exploitation</i>	10	0	33.066

The keyness analysis for SDG 8 shows that several keywords related to employment, labor conditions, and inclusive economic development appear significantly more frequently in the UGM news corpus compared to the Elsevier SDG keyword reference list. The highest keyness value is found for *tourism* (143.599), indicating that UGM discourse frequently frames economic growth through tourism development, likely reflecting both Indonesia's national economic priorities and the university's research, community service, and policy engagement activities in regions where tourism is a major livelihood sector. In addition to the recurrent use of *tourism* (48), there are some recurrent uses of keywords like *growth* (41), and *trade* (23) which suggests tourism and trade-driven development discourses reflecting macroeconomic orientation which is Target 8.1 (Sustained Economic Growth).

The repeated presence of the terms such as *labour*, *employment*, *labor*, *wage*, *productivity*, and *technological* further suggest a strong focus on work-related conditions and workforce development, as well as research-led economic modernization aligning with SDG 8's emphasis on decent work, fair compensation, and capacity building. This is also closely related to the SDG Target 8.2 (Higher Levels of Productivity and Innovation) and 8.5 (Decent Work and Equal Pay).

In addition, the recurrent use of *child*, *children*, *women*, *trafficking*, *slavery*, and *exploitation* indicates discourse attention to vulnerable groups and labor rights protections. These terms reflect a human rights dimension within UGM's SDG 8 communication, positioning the institution as not only promoting economic productivity but also

addressing inequalities and risks of exploitation. This is closely related to the SDG Target 8.7 (Eradication of Forced Labor and Human Trafficking) and Target 8.8 (Worker Rights and Safe Working Environments).

Meanwhile, *growth*, *trade*, *credit*, and *income* highlight inclusive economic mechanisms associated with local and regional development, whereas *technological*, *policy*, and *economic* suggest that innovation and governance serve as the frame through which economic issues are discussed. This is closely related to target 8.3 (MSMEs and Entrepreneurship Support).

Overall, the SDG 8 keyword profile demonstrates a dual narrative: promoting sustainable economic development while concurrently advocating for socially responsible labor practices and protections for marginalized populations.

### 3.3 Findings on SDG 9: Industry, Innovation and Infrastructure

The keyness statistics for SDG 9 indicate 132/618 keyword types and 2,023/3,099 keyword tokens. This means that there are 132 unique lexical items in the SDG 9 corpus that are identified as keywords relative to the reference corpus. This is out of a total of 618 word types in the SDG 9 dataset. In other words, approximately 21.36% of all unique words in the SDG 9 corpus are considered statistically significant keywords. This suggests a relatively high lexical specificity, meaning that the SDG 9 texts have a focused vocabulary that distinctly characterizes discussions related to industry, innovation, and infrastructure

In addition, the value of 2,023/3,099 keyword tokens mean that the keywords appear 2,023 times out of 3,099 total word occurrences in the SDG 9 corpus, indicating that about 65.28% of the total words used in the SDG 9 dataset are instances of these statistically significant keywords. The high proportion of keyword tokens suggests that the discourse is dense with topic-relevant terms, reinforcing the thematic strength and concentration of SDG 9-related language in the corpus.

**Table 3.** Log-likelihood values of Top 20 SDG 9 keywords (UGM vs. Reference Corpus)

No	Keyword	Frequency in UGM Corpus	Frequency in Reference Corpus	Log-likelihood
1	<i>manufacturing</i>	80	4	361.586
2	<i>fund</i>	70	0	343.7
3	<i>access</i>	67	13	260.215
4	<i>industrial</i>	64	18	231.017
5	<i>energy</i>	58	13	219.29
6	<i>infrastructures</i>	45	1	211.158
7	<i>small</i>	45	6	184.743
8	<i>infrastructure</i>	69	55	178.232
9	<i>medium</i>	44	7	176.157
10	<i>sized</i>	32	1	147.975
11	<i>industries</i>	38	12	133.262
12	<i>machining</i>	25	0	122.415
13	<i>internet</i>	32	7	121.308
14	<i>emission</i>	25	1	114.118
15	<i>employment</i>	23	0	112.608

16	<i>industry</i>	56	72	112.142
17	<i>mobile</i>	24	3	99.215
18	<i>waste</i>	27	7	98.91
19	<i>micro</i>	24	4	95.265
20	<i>reduction</i>	22	2	94.298

The log-likelihood finding as shown in Table 3 above highlights the top twenty keywords with the highest statistical significance, including *manufacturing*, *industrial*, *energy*, *infrastructure*, *small*, *medium*, *industries*, and *employment*. All show significant higher log-likelihood values compared to the reference corpus. These keywords highlight the university's focus on strengthening innovation ecosystems, industrial resilience, and infrastructure development as central themes of SDG 9, particularly Target 9.2. (Promote Inclusive and Sustainable Industrialization).

The high log-likelihood values indicate that these terms appear more frequently and meaningfully within UGM's institutional discourse. The term *manufacturing* shows the strongest keyness (LL = 361.586), highlighting a discourse focus on industrial production, applied research, and value-added technology. This suggests that UGM frequently positions itself in relation to national industrial upgrading efforts, which aligns with SDG 9's emphasis on building resilient industry and promoting innovation. The prominence of *fund* (LL = 343.700) suggests strong discursive attention to financing mechanisms, research investments, and funding support. This resonates with SDG 9's target on increasing access to financial services and support structures for innovation and enterprise development [16]. In Indonesia, access to funding has been recognized as a central constraint in scaling micro, small, and medium-sized industries [5], making this lexical prominence contextually relevant.

Keywords such as *small*, *medium*, *sized*, *industries*, *micro*, and *employment* collectively indicate attention to the MSME sector, which is central to Indonesia's economic structure and labor absorption. These keywords align with Target 9.3, encouraging the integration of small-scale enterprises into value chains and markets. MSMEs account for more than 90% of national employment and play a key role in innovation diffusion [23]. The linguistic patterns therefore reflect UGM's role in promoting capacity building, training, and empowerment efforts targeted at MSMEs—consistent with national development priorities.

Similarly, the keywords *industrial*, *infrastructure*, and *infrastructures* (LL range ≈ 178–231) reveal a discourse centered on the development and upgrading of physical and technological infrastructure. These keywords correspond to Target 9.1, emphasizing the development of reliable, sustainable infrastructure to support economic growth and human well-being [14].

Infrastructure is widely recognized as a foundational driver of industrial competitiveness and regional innovation ecosystems [10]. The presence of *internet* further suggests emphasis on digital infrastructure, aligning with the growing policy narrative on accelerating digital transformation in higher education and the economy. Furthermore, the recurrent use of terms such as *energy*, *internet*, and *mobile* connect to Target 9.4 and 9.C, which advocate upgrading industries and infrastructures to make them sustainable and expanding access to information and communication technologies.

Environmental sustainability appears in the keywords such as *emission*, *waste*, and *reduction*, yet these terms have lower keyness rankings relative to industrial and economic terms. This suggests that while sustainability is present in discourse, industrial competitiveness and innovation functions are prioritized more strongly than ecological dimensions. Such prioritization is consistent with findings that institutional SDG communication often highlights areas of strategic advantage or reputational value [26]. The inclusion of *machining* and *emission* indicates UGM's engagement with environmentally responsible industrial processes, aligning with Target 9.4 on clean technologies and resource efficiency.

Overall, the findings indicate that UGM's communication of SDG 9 highlights innovation-led industrial growth, MSME empowerment, funding access, and infrastructure enhancement, positioning the university as the key actor in national industrial modernization. Environmental concerns are acknowledged but are positioned as complementary rather than central to the discourse.

### 3.4 Findings on SDG 17: Partnership for the Goals

From the comparison done between the UGM SDG 17 articles as the target corpus and the Elsevier keywords of SDG 17 as the reference corpus, the target corpus has 63 unique keywords of SDG 17 that stand out statistically in the UGM corpus when compared to 787 unique word types in total, in the reference corpus. This means about 8% of all unique words in the target corpus are keywords which suggests that SDG 17 discourse in UGM news is selective, using a focused set of collaboration and partnership-related vocabulary, rather than broad lexical variety. Also, the target corpus has 447 keyword tokens, compared to 1,492 total occurrences in the reference corpus. This means that SDG 17 associated keywords make up about 29.95% of all keyword instances. Thus, even though SDG 17 uses a relatively small variety of specialized vocabulary (63 types), that vocabulary is used frequently and repeatedly across the corpus.

**Table 4.** Log-likelihood values of Top 20 SDG 17 keywords (UGM vs. Reference Corpus)

No	Keyword	Frequency in UGM Corpus	Frequency in Reference Corpus	Log-likelihood
1	<i>trade</i>	32	3	163.81
2	<i>debt</i>	24	0	137.833
3	<i>tax</i>	22	0	126.318
4	<i>south</i>	20	10	77.778
5	<i>trading</i>	14	1	73.082
6	<i>investment</i>	12	3	54.162
7	<i>country</i>	9	0	51.601
8	<i>capacity</i>	8	1	39.701
9	<i>implementation</i>	10	5	38.829
10	<i>agreement</i>	11	8	38.151
11	<i>foreign</i>	8	2	36.088
12	<i>fiscal</i>	6	0	34.389
13	<i>risk</i>	7	1	34.214
14	<i>state</i>	10	8	33.544
15	<i>co</i>	7	2	30.825

16	<i>use</i>	7	2	30.825
17	<i>sovereign</i>	6	1	28.765
18	<i>regime</i>	5	0	28.655
19	<i>collab</i>	5	0	28.655
20	<i>capability</i>	5	0	28.655

The keyness analysis generated 63 keyword types and 477 keyword tokens associated with SDG 17 within the UGM institutional news corpus. These results indicate that the discourse surrounding SDG 17 is characterized by a relatively concentrated but frequently repeated lexical set, suggesting an intentional thematic focus rather than lexical diversity. This aligns with corpus linguistic perspectives that note how institutional discourse often relies on recurrent, formulaic lexical patterns to foreground strategic identity positioning [3; 20]. The high frequency of repeated partnership-related keywords demonstrates the centrality of collaborative engagement within UGM's framing of sustainability initiatives. This finding demonstrates that UGM's SDG-related publications emphasize international and institutional partnerships that contribute to achieving SDG 17.

The prominence of *trade*, *investment*, and *agreement* aligns with target 17.10 and 17.12, which aim to promote equitable trading systems and increase exports from developing countries. The high occurrence of *debt* and *tax* corresponds to target 17.1 and Target 17.4, focusing on strengthening domestic resource mobilization and ensuring sustainable debt management [28]. Meanwhile, keywords such as *capacity* and *implementation* reflect Targets 17.6-17.9, which emphasize technology transfer, capacity-building, and policy coherence across nations.

These lexical patterns suggest that UGM's engagement with SDG 17 centers on discourses of economic cooperation, fiscal governance, and transnational collaboration. The repeated use of *country*, *foreign*, and *agreement* further signals attention to global partnership mechanisms and diplomatic collaboration between Indonesia and other nations, consistent with studies showing that universities play a critical role in international policy networks supporting the SDGs [22].

Overall, the keyness results confirm that SDG 17-related discourse at UGM prioritizes systemic collaboration and institutional partnerships—core drivers for sustainable development—thereby demonstrating strong alignment with global goals to strengthen the means of implementation and revitalize partnerships.

From the findings above, the keywords with the highest log-likelihood values are primarily associated with international cooperation, capacity building, inter-organizational collaboration, and formal alliance-building. The recurrent presence of these terms signals that partnerships are not represented merely as supportive activities, but as core operational mechanisms through which sustainability objectives are pursued. This discursive pattern reflects the intent of SDG 17, which emphasizes strengthening the *means of implementation* and revitalizing global partnerships to support sustainable development outcomes [27]. In UGM's communication, partnerships are consistently positioned as strategic assets for research collaboration, community development, policy-making influence, and institutional networking.

Furthermore, several keywords relate to resource mobilization, governance structures, and multi-stakeholder coordination, indicating that the partnerships described are typically formalized and structured rather than informal or ad hoc. This suggests that

UGM frames partnerships as institutional commitments embedded in negotiated agreements and sustained cooperation frameworks. Such discourse is reflective of broader trends in higher education internationalization, where universities increasingly define their global roles through collaborative engagement and network participation [13].

Overall, the findings illustrate that UGM's SDG 17 discourse constructs the university as an active and globally engaged institutional actor, continually working with governmental agencies, private-sector partners, and international academic institutions. Partnerships are presented as instrumental pathways for achieving broader sustainable development outcomes, while simultaneously reinforcing UGM's identity as a key contributor within regional and global development ecosystems. The repetition and prominence of partnership-related vocabulary thus underscores SDG 17's function as a strategic cornerstone in UGM's sustainability communication practices.

## 4 Conclusion

This study mapped the keywords and thematic orientations of Universitas Gadjah Mada's (UGM) SDG-related publications by employing a corpus-based log-likelihood keyness analysis to compare the lexical prominence of terms in UGM's corpus with Elsevier's SDG keyword list. The analysis of SDGs 7, 8, 9, and 17 reveals clear thematic distinctions that reflect UGM's strategic contributions to the UN 2030 Agenda for Sustainable Development.

The lexical mapping for SDG 7 (Affordable and Clean Energy) demonstrates high frequencies of *energy*, *solar*, *wind*, and *hydro*, emphasizing the university's research strength in renewable energy innovation and sustainable power systems. SDG 8 (Decent Work and Economic Growth) highlights *tourism*, *labour*, *employment*, and *productivity*, showing UGM's focus on inclusive growth, decent employment, and human capital development. The findings for SDG 9 (Industry, Innovation and Infrastructure) reveal prominent terms such as *manufacturing*, *industrial*, *infrastructure*, *innovation*, and *technology*, signifying strong engagement in industrial transformation, technological capacity building, and infrastructural advancement. Lastly, SDG 17 (Partnerships for the Goals) features key lexical items such as *trade*, *debt*, *tax*, *investment*, and *agreement*, which underscore the institution's active participation in international cooperation, policy partnerships, and financial collaboration to support global sustainability efforts.

Overall, the keyness results indicate that UGM's SDG-related discourse integrates three major thematic dimensions: (1) technological and energy innovation, (2) inclusive economic development, and (3) global partnerships and institutional collaboration. These findings affirm the university's pivotal role in promoting sustainable development through research communication and partnership-based initiatives. Future studies could extend this work by incorporating longitudinal data across multiple years, comparing institutional corpora among ASEAN universities, and employing semantic network or topic modelling approaches to capture deeper thematic relationships among SDG-related discourses.

**Acknowledgments.** This study was funded by Vocational School Universitas Gadjah Mada under grant number 152/UN1/SV/K/2024.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Anthony, L. AntConc (Version 4.2.0) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/> (2022)
2. Baker, P. *Using Corpora in Discourse Analysis*. London: Routledge. (2006)
3. Baker, P. *Using corpora in discourse analysis*. Continuum. (2006)
4. Baker, P. *Sociolinguistics and corpus linguistics*. Edinburgh University Press. (2010)
5. Bappenas. *Rencana Pembangunan Jangka Menengah Nasional 2020–2024*. Kementerian PPN/Bappenas. (2020)
6. Bedard-Vallee, A., James, C., & Roberge, G. Elsevier 2023 Sustainable Development Goals (SDGs) Mapping [Dataset, version 1]. Elsevier BV (2023). <https://doi.org/10.17632/y2zyy9vwzy.1>
7. Ceulemans, K., Molderez, I., & Van Liedekerke, L. Sustainability reporting in higher education: A comprehensive review of the recent literature and paths for further research. *Journal of Cleaner Production*, 106, 127–143 (2015). <https://doi.org/10.1016/j.jclepro.2014.09.052>
8. Cui, C. A three-dimensional discourse analysis of SDGs reports and Apple's environmental progress reports: A corpus-based study. In *Proceedings of the 3rd International Conference on Culture, Design and Social Development (CDSO 2023)*. pp. 229–234. Atlantic Press (2024). [https://doi.org/10.2991/978-2-38476-222-4\\_27](https://doi.org/10.2991/978-2-38476-222-4_27)
9. Fauzi, M. A., Abdul Rahman, A. R., & Lee, C. K. A systematic bibliometric review of the United Nation's SDGs: Which are the most related to higher education institutions? *International Journal of Sustainability in Higher Education*, 24(3), 637–659 (2022). <http://dx.doi.org/10.1108/IJSHE-12-2021-0520>
10. Heras-Saizarbitoria, I., Sáez, L., & Boiral, O. Organizational communication and the Sustainable Development Goals: A systematic review. *Journal of Business Ethics*, 180(4), 885–903. (2022)
11. Hou, M., & Ayers, D. F. Community college discourses of sustainability in an academic and professional corpus. *International Journal of Sustainability in Higher Education*, 24(8), 1669–1686 (2023). <https://doi.org/10.1108/IJSHE-04-2022-0113>
12. Kilgarriff, A. Simple maths for keywords. In *Proceedings of the Corpus Linguistics Conference*. University of Liverpool. (2009)

13. Knight, J. Concepts, rationales, and interpretive frameworks in the internationalization of higher education. In D. Deardorff, H. de Wit, J. Heyl, & T. Adams (Eds.), *The SAGE Handbook of International Higher Education*. pp. 27–42. SAGE, Thousand Oaks. (2012)
14. Lall, S. Infrastructure and industrial competitiveness: A structural economics perspective. *Oxford Development Studies*, **48**(2), 145–160. (2020)
15. Nikolaou, I. I., Tsalis, T. A., Trevelopoulos, N. S., et al. Exploring the sustainable reporting practices of universities in relation to the United Nations' 2030 Agenda for sustainable development. *Discover Sustainability*, **4**, 46 (2023). <https://doi.org/10.1007/s43621-023-00167-1>
16. OECD. *Financing SMEs and entrepreneurs 2021: An OECD scoreboard*. OECD Publishing. (2021)
17. Rayson, P., & Garside, R. Comparing corpora using frequency profiling. In *Proceedings of the ACL Workshop on Comparing Corpora*. pp. 1–6 (2000)
18. Satuan Penjaminan Mutu dan Reputasi Universitas Gadjah Mada. *TCK dan Program Menuju UGM Sustainable Campus [Performance Achievement Targets and Programs Towards UGM Sustainable Campus] [PowerPoint presentation]* (2024)
19. Scott, M. Problems in investigating keyness or clearing the undergrowth and marking out trails. *International Journal of Corpus Linguistics*, **15**(1), 1–19 (2010)
20. Scott, M., & Tribble, C. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins. (2006)
21. Smeuninx, N., De Clerck, B., & Aerts, W. Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP. *International Journal of Business Communication*, **57**(1), 52–85 (2020). <https://doi.org/10.1177/2329488416675456>
22. Stafford-Smith, M., Griggs, D., Gaffney, O., et al. Integration: The key to implementing the Sustainable Development Goals. *Sustainability Science*, **12**, 911–919 (2017). <https://doi.org/10.1007/s11625-016-0383-3>
23. Tambunan, T. MSMEs in Indonesia: Important problems and constraints. *Journal of Southeast Asian Economies*, **36**(3), 1–19. (2019)
24. *Times Higher Education. Report: How university partnerships are accelerating progress towards the SDGs*. <https://www.timeshighereducation.com/digital-editions/report-how-university-partnerships-are-accelerating-progress-towards-sdgs> (2022)
25. United Nations. *The sustainable development agenda [Retired page]*. <https://www.un.org/sustainabledevelopment/development-agenda-retired/> (2018)
26. United Nations. *The Sustainable Development Goals Report 2020*. UN, New York. (2020)
27. United Nations General Assembly. New York. (2015)
28. United Nations Statistics Division. Indicator 17.10.1: Worldwide weighted tariff-average [Metadata for Sustainable Development Goal 17, Target 17.10] (2025). <https://unstats.un.org/sdgs/metadata/?Goal=17&Target=17.10>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

