Sustainable Development Goals (SDGs) Interlinkages Analysis Based on Text Mining

Dr. Hassan Rabie *

Abstract

5 years passed and only 10 years left to achieve Sustainable Development Goals (SDGs), which were seen as an indivisible agenda. Since the launching of SDGs and the work on understanding and analyzing the nature of SDG interlinkages is going, which is vital to break down vertical siloes and to support decision-making process. Lack of integration across various sectors may result in incoherent policies and has been one of the major hindrances sustainable development approaches previously taken. Several methodological approaches for analyzing SDGs interlinkages have been proposed. The linguistic approach is one of the main approaches, in which, the assessment of the interlinkages between SDGs is based on their respective wording such as a keyword search. Once obtained and identified, the usage of such (clear) keywords make no confusion about the existence of interlinkages. Therefore, the assumption, two goals are interlinked if they have at least one thematic area keywords in common. The linguistic approach is applied using traditional methods, which lead to time consuming and inaccurate search process. This paper reviews the previous SDGs interlinkages methods; focusing on the linguistic approach. Also, a new text mining approach has been proposed to identify the SDGs interlinkages. The new text mining method could offer a signpost as a new method to improve process of identifying the SDGs interlinkages. The MICMAC analysis has been applied to further investigate for the SDGs interlinkages and to highlight the interrelationships between SDGs.

Keywords: Text Mining, Sustainable Development, SDGs Interlinkages, linguistic Approach MICMAC.

ملخص

منذ إطلاق أهداف التنمية المستدامة والعمل جارى على فهم وتحليل طبيعة الروابط بين أهداف التنمية المستدامة. قد يؤدي عدم التكامل عبر مختلف القطاعات إلى سياسات غير متماسكة وهو الأمر الذي شكل أحد العوائق الرئيسة التي واجهت نهج التنمية المستدامة في السنوات السابقة .

تستعرض هذه الورقة أساليب تحليل الروابط بين أهداف التنمية المستدامة؛ مع التركيز على النهج اللغوي، وتقترح نهج جديد للتنقيب عن النصوص لتحديد الروابط بين أهداف التنمية المستدامة، حيث يمكن لطريقة التنقيب عن النصوص أن تشكل مدخلاً جديداً لتحسين عملية تحديد وتحليل الروابط البينية بين أهداف التنمية المستدامة. وأيضاً تم تطبيق تحليل MICMAC لمزيد من تعميق البحث عن طبيعة وأبعاد الترابطات، وتسليط الضوء على العلاقات المتبادلة بين أهداف التنمية المستدامة بعضها البعض.

الكلمات الدالة: تعدين النصوص – التنمية المستدامة – الروابط بين أهداف التنمية المستدامة – المنهج اللغوي – MICMAC

* **Lecturer at Planning Techniques Center – Institute of National Planning.**

1. Introduction

The former UN Secretary General Ban Ki-moon mentioned that, SDGs will provide a road map that builds on the Millennium Development Goals (MDGs) (UN, 2018). For some countries, governments prioritized SDGs to guide national developments plans and policy measures (Zhou & Moinuddin, 2017).

Sustainable development is defined as "sustainable the development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Keeble & war, 1988). In 2015, the United Nations (UN) Member adopted the Sustainable Development Goals (SDGs) as a universal call to end hunger, protect the environment and planet, and ensure that all people enjoy peace by 2030. They agreed to launch a process to develop a set of sustainable development goals (SDGs). The goals should address in a balanced way all the three dimensions of sustainable development (Economic, Social, and Environmental) and should be coherent and integrated. The Open Working Group (OWG) proposed a comprehensive set of 17 goals and 169 targets, to deal with the problems affect developing and developed countries.

SDGs are inter-connected and can't be discussed separately (Miola, Borchardt, Neher, & Buscaglia, 2019).Many interlinkages can exist among the SDGs. Also, many literatures have mentioned that, SDGs were chastised for having too many objectives and not having a clear sense of priorities, since the interlinkages between SDGs are very complex (Elder, 2016). During the SDG negotiations this complexity was probably not feasible for many governments and researches to address, but now, SDGs are in the implementation stage, it is necessary to map these out (Elder, 2016).

Therefore, we need a systemic approach to be adopted to deal with the aspect of interlinkages. Most of the strategic plans and decisions to tackle and achieve SDGs are focused on one goal, or a small set of SDGs, since most of decision-makers are siloed and fragmented. Governments may also prioritize small subsets of SDGs to meet national objectives (Scharlemann et al., 2020). Since some decisions and actions to achieve any SDG most probably affect the accomplishment of the other SDGs, therefore, a better studying of the interactions and interlinkages between SDGs is urgently needed (Scharlemann et al., 2020).

Text Mining (TM) describes a range of methods for analyzing and treating semi-structured and unstructured text data (Miner et al., 2012). The main objective of TM is to discover the unknown, useful and hidden knowledge from the semi-structured and unstructured text data (Ding, Li, & Fan, 2018). According to (Kwartler, 2017), TM is the process of extracting actionable insights from unstructured text. Essentially, it means going from an unorganized state to a summarized and structured state. In text mining, the text will be processed and transformed into a numerical representation. Text mining is the process of taking unorganized sources of text, and applying standardized analytical steps, resulting in a concise insight and recommendation.

Since the launching of SDGs and the work on understanding and analyzing the nature of SDG interlinkages is going. Many approaches and methods have been proposed and developed. However, these approaches while studying and searching for interlinkages between SDGs depend on using traditional and/or manual methods, such methods are time consuming, and labor-intensive process, unsystematic and might be subjective.

This paper proposes a new text mining approach to identify SDGs interlinkages. The approach supports the identification of interlinkages in a systemic way through studying and analyzing the intersection among the keywords of the SDGs text. The objective is to understand the relationships and to identify the SDGs interlinkages, so that decision-makers can avoid the silo model of policymakers. Up to the knowledge of the author, no existing text mining approach for identifying the SDGs interlinkages has been proposed or applied in the literature.

2. Literature Review

The 2030 Agenda has noted that SDGs – which address diverse issues – are integrated and indivisible, however the agenda has not elaborated how they are interlinked. Interlinkages are a key aspect of integration. SDGs address specific issues associated to the sustainable development; the goals rely on each other. However, policymakers, planners and decision-makers work in silos without sufficiently coordinating or cooperating with each other, and they do not have the means to identify and prioritize reinforcing or trade-off (Zhou & Moinuddin, 2017). Lack of integration across various sectors may result in incoherent policies and has been one of the major hindrances sustainable development approaches previously taken (Le Blanc, 2015). Highlighting and understanding SDG interlinkages is critical to break down vertical siloes (Cutter, Osborn, Romano, & Ullah, 2015).

The first work related to the SDGs interlinkages was done by OWG, which released nineteen focus areas for the post-2015 seventeen global goals, through publishing an annex document where the interlinkages among 19 focus areas were listed up (OWG, 2014). The list shows that most of the focus areas tend to have varying degrees of interlinkages with other areas (Zhou & Moinuddin, 2017). Since then, many studies documenting the SDG interlinkages have adopted different approaches and methodologies. The integrated nature of the SDGs is studied by several methods; however, many researches have approached and studied the interlinkages between SDGs by using the text/wording of the SDGs through matching goals and/or targets with each other – Linguistic approach – (Miola et al., 2019). Such that, two goals are interlinked if they have in common some explicit text or keyword.

(Cutter et al., 2015) reported the interlinkages among the sustainable development goals – based on a draft version. They proposed the first linguistic approach to identify and list the SDGs interlinkages by focusing on SDGs, where the goal wording used to link other SDGs. (Le Blanc, 2015) presented a systemic approach by seeing the SDGs – except Goal 17– as a graph, in which links among SDGS established through targets that refer to multiple SDGs. The foundation for the interlinkages is a matrix that connects every SDG target to all of the SDGs to which its wording refers; i.e., based on the context of SDG wording.

In a further elaboration, (Vladimirova & Le Blanc, 2016) identified links between Goal 4 “Education” and other SDGs, based on a comprehensive content analysis of thirty seven reports. They manually examined the connections (interlinkages) between education and other SDGs, with manual individual analysis of each occurrence of word “Education.” Statements that contained conclusions and analytical were extracted and formed the basic corpus. Statements referring “Education” and the 16 other SDGs were considered. Statements grouped into four analytical categories, then, analyzed according to the causal direction of the links.

(Nunes, Lee, & O'Riordan, 2016) proposed a linguistic framework to connect the health and non-health sectors by identifying to which sectors the goals' text are referring to. They concluded that health and non-health goals are cross-cutting and unifying themes. To understand the nature of interlinkages in SDGs (Coopman, Osborn, Ullah, Auckland, & Long, 2016) proposed a linguistic methodology to analyze and identify different types of interlinkage, focusing particularly on the interlinkages between the targets in Goal 12 and targets in other SDGs, by the wording of the goals. Based on a list of keywords, (Le Blanc, Freire, & Vierros, 2017) maps the interlinkages among targets of SDGs dedicated to Goal 14 "Oceans". Mainly, the authors analyzed 124 reports from the UNs system and other international organizations using traditional and manual methods.

In these researches, the interlinkages identified using traditional and/or manual search methods to find keywords in the SDGs. Such work is very difficult, time consuming, labor-intensive process, unsystematic and might be subjective. Text mining does all this without forcing an individual to read the entire corpus. Therefore, a systemic text mining approach needs to be developed to identify SDG interlinkages.

The studies that focus on analyzing the SDGs text to identify SDG interlinkages are still limited. This study would like to fill this gap based on the objective analysis of the all the SDGs text in English. As the corpus of SDGs is not standardized (unstructured), TM approach must be used to discover the hidden SDGs interlinkages therein. The key idea is that text mining approaches can be used to analyze SDGs corpus to identify SDG interlinkages.

Text mining approaches have already been used in the sustainable development literature. (Rivera, Minsker, Work, & Roth, 2014) proposed a text mining framework to assess and track sustainability indicators. TM approach was used to analyze sustainability content in news articles. The results were compared with sustainability indicators reports created by Sustainable San Mateo County (SSMC) using traditional methods. The findings revealed that the text mining approach was able to identify all of the indicators indicated as essential in the reports, and that the technology has the capacity to uncover recurrent patterns.

(Fabbrizzi et al., 2016) studied the impact of the Rio Summit in the literature of sustainability. TM approaches were combined for the investigation of scientific researches. The findings reveal that in the pre-Rio period, agricultural ecology and environmental components were dominating, and that after then, the research was expanded to include other issues such as energy, urban regions, and the importance of indicators. (Shin et al., 2018) examined the sustainability of maritime by using TM approach. Using a text mining methodology, the authors produced an analysis of major journal publications in the field of marine studies published between 1993 and 2017.

(Bakhtin, Khabirova, Kuzminov, & Thurner, 2020) presented a text mining study to identify the emerging topics in food and agriculture production. They presented analyzed unstructured text data sources by applying innovative TM approach. They conclude that the findings can be applied to any organization in the agriculture and food sector's strategic planning process. (Sebestyén, Domokos, & Abonyi, 2020) used TM approach to analyze the reviews of the tailored strategies for the achievement of Sustainable Development Goals for 75 countries. TM results show that Goal 5 "Gender equality" is the most studied goal. The Goal 8 "decent work and economic growth" is the second most mentioned goal, while Goal 1 "End poverty" is the least focused goal.

3. Research Methodology

3.1 Research Outline

In this paper, text mining and analysis methods are applied to disclose the main features of SDGs. Text mining is the process of deriving information from text data which is not previously known and not easy to be revealed (Zhang, Fleyeh, Wang, & Lu, 2019).
Figure 1 shows the research outline. The first step is data collection, which explains how text data for the SDGs was gathered. Data preprocessing is the second step, which tries to enhance the quality of the collected data in preparation for future mining and analysis. Finally, two text mining algorithms are used in the knowledge discovery process to derive meaningful insights from SDG text data.

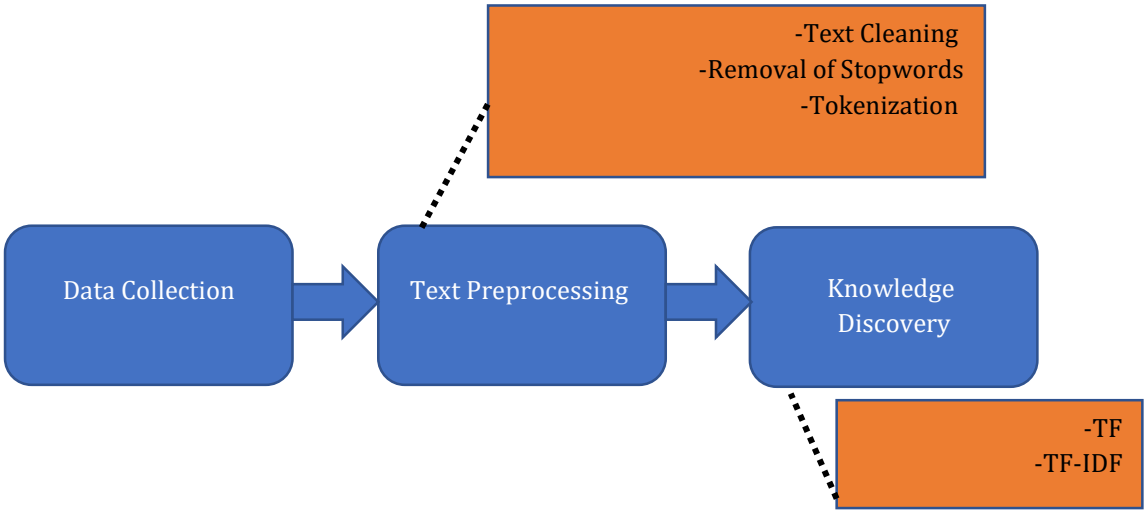

Figure 1: Research Outline

3.2 Data Collection and Text Preprocessing

The text of SDGs (goals and their related targets) is used as text inputs for the text mining process. All SDGs have been included in the analysis except goal 17, which is considered an enabling goal and supporting for other SDGs (Le Blanc, 2015; Scharlemann et al., 2020). The official text includes the global goals, and targets framework as contained in A/RES/71/313[1]. The corpus contains about 9,000 words.
SDGs text data has downloaded and stored in a "CSV" file. Text preprocessing is used to transform unstructured text into structured text in order to improve the quality of text data for further analysis. In this paper, TM analysis was executed using the R programming languages (Team, 2018). In R, different set of packages were used to perform the analysis such as "tiditext", "tidyverse", "stringr", "dplyr", "wordcloud" and "tidyr". Text preprocessing involves:

1. Convert all letters to lower case, remove Punctuation Marks, remove Numbers, and strip Whitespace from the text of the goals (Zhang et al., 2019).

[1] https://unstats.un.org/sdgs/indicators/indicators-list/

2. Remove Stopwords from the text, since some words are of little use for data analysis, and not all words are important, such as (“the”, “an”, “for”, and “in”), therefore, these words were removed from the text data (Galati & Bigliardi, 2019; Zhang et al., 2019).
3. Finally, transform unstructured text data into a two-dimensional data table with one word/token. This process called tokenization where the text is split into individual tokens (Silge & Robinson, 2017; Zhang et al., 2019).

3.3 Knowledge Discovery

The absence of a general methodology to perform text mining has become a great challenge (Buenaño-Fernandez, González, Gil, & Luján-Mora, 2020). Bag-of-words is one of the most common and useful approaches of text mining (Siegel, 2013). Bag-of-words easy to analyze and understand and not computationally expensive (Jones-Diette, Dean, Cobb, & Brennan, 2019). Usually, the process starts by converting unstructured text into a structured format then applying a group of approaches and algorithms to extract patterns and uncover new knowledge (Kotu & Deshpande, 2019; Miner et al., 2012). In this paper, two Bag-of-words approaches were, (1) Analyzing Term Frequency (TF) and (2) Analyzing most term frequently using TF-IDF. Although, these approaches are the most straightforward approaches to analyzing and mining SDGs text data. It can quickly identify the key characteristics and interlinkages of SDGs.

4. Results and Discussion

4.1 Term Frequently

To understand more about SDGs; an investigation of the words which was used in describing them will be presented. This can be done by finding the most commonly used words/tokens in the SDGs. Word frequency, which also called term frequency (TF), is an easy way to measure how a word is used (Silge & Robinson, 2017).

Term frequency is the common method and straightforward approach to analyze text data. Wordcloud is a common visual representation of word frequency of text data and a quick way to summarize large amounts of text data. SDGs wordcloud shows that the most frequently words frequently mentioned in the SDGs text (goals and targets). Figure 2 shows words with larger font sizes indicate more frequently occurring terms. Figure 2 indicates the top-5 frequently used words in the SDG text are: Countries, Sustainable, Including, Developing, and Development. This means SDGs is focusing on the sustainability for developing and developed countries.

Figure 2: SDGs Worldcloud

4.2 Term Frequently–Inverse Document Frequency (TF-IDF)

To achieve the purpose of this paper, a new approach based on the TF-IDF is implemented. To identify the SDG interlinkages, text mining approach adopted in this paper does not rely on predetermined notions. Most TF may appear in a small number of SDGs and may be treated to be more important than terms and words occurring in every goal separately. Therefore, other measure TF-IDF used to measure the importance of each word in each goal.

TF-IDF used to identify the relative importance of each word within each SDG and also can be used to quantify what a goal is about (Park & Kremer, 2017). TF-IDF is easy to compute and understand, and is considered one of the most important TM methods to show the relative importance of a word. TF-IDF is used to measure the relative importance of each word in different goals and to highlight terms that are not frequently used but have unique representative power in each goal (Ding et al., 2018; Zhang et al., 2019). Therefore, TF-IDF could be used to generate SDGs keywords and also to summarize any corpus such as SDGs using a few keywords, by computing the importance of each word within each goal.

4.2.1 TF-IDF to identify SDGs interlinkages

As mentioned earlier, most of linguist approach for the SDGs were performed using traditional or manual methods. In this paper, the basis for identifying SDGs interlinkages is a matrix that links every goal to all other SDGs to which its wording (keywords) refers i.e., have in common at least one thematic area keyword, such as Poverty, Water, Economic …etc. Therefore, and for SDGs interlinkages, TF-IDF is applied to identify these thematic area keywords. For the linguistic approach assesse of the interlinkages between SDGs based on their respective

wording such as a keyword. The linguist approach concept, two goals are interlinked if they have in common at least one word (Le Blanc, 2015; Miola et al., 2019).

The proposed text mining approach is depending on using TF-IDF method which is a word statistic method for text feature extraction, i.e. "keywords". The approach can be applied as follows:

1. Find and remove words that have been mentioned on one goal, since such these words have no importance in identifying the interlinkages between any two goals – for the purpose of this paper–.
2. Calculate the TF-IDF for every word in every goal using Tidytext, R packages. Each word has more than TF-IDF value, since each word has different importance in each goal.
3. Sum up the values of TF-IDF for each word and then sort the list in a descending order. Now, we have a sorted long list with more than 320 words, ranked form the largest to the smallest values of TF-IDF.
4. To obtain meaningful interlinkages between SDGs, two filters have been applied on the long list:
   a. Deleting words that have no thematic area meaning and common words, such as "change", "forms", "modern", "affordable", "impacts", "combat", "reduce" "sustainable", and "resources". To make sure of the list of thematic area keywords, a comparison with other common lists have been hold. These keyword lists are: (1) keywords listed in the sustainability science in the global landscape report[2], (2) Compiled Keywords for SDG Mapping[3].
   b. Selecting keywords in a high TF-IDF position in the rank, therefore, words in the first and second quartile are considered and exclude words in the third or fourth quartile as in (Galati & Bigliardi, 2019).

4.2.2 Keywords-SDGs Matrix

Table 1 shows the Keywords-SDGs links matrix. The first column contains the 37 thematic area keywords, which have been obtained after applying the four steps in the previous proposed approach. The next columns of Table 1 show all the SDGs. The intersection of rows and columns (cells) contains "√" if the keyword is founded in the text of goal at least one time.

[2] https://www.elsevier.com/__data/assets/pdf_file/0018/119061/SustainabilityScienceReport-Web.pdf

[3] http://ap-unsdsn.org/webinar-mapping-university-contributions-to-the-sdgs/compiled-keywords-for-sdg-mapping_final_17-05-10/

Table 1: Keywords-SDGs links Matrix

Keyword	SDGs																Count
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Women	√	√		√	√	√		√			√		√				**8**
Technology	√	√		√	√		√		√					√			**7**
Economic	√				√			√	√	√	√			√			**7**
Land	√	√			√		√							√	√		**6**
Children	√	√	√	√							√					√	**6**
Water			√			√					√	√			√		**5**
Vulnerable	√	√		√		√					√						**5**
Trade		√	√					√		√				√			**5**
Climate	√	√									√		√				**4**
Health			√		√							√		√			**4**
Education			√	√				√					√				**4**
Infrastructure		√			√		√		√								**4**
Girls		√		√	√	√											**4**
Poor	√	√									√	√					**4**
Ecosystems		√				√								√	√		**4**
Social	√				√					√	√						**4**
Disasters	√	√									√		√				**4**
Employment		√		√				√	√								**4**
Trafficking					√			√							√	√	**4**
Environmental	√							√			√	√					**4**
Production		√						√				√					**3**
Gender	√			√	√												**3**
Violence				√	√											√	**3**
Youth				√				√					√				**3**
Equality				√	√					√							**3**
Child				√	√			√									**3**
Pollution			√			√								√			**3**
Food		√										√					**2**
Poverty	√														√		**2**
Biodiversity														√	√		**2**
Consumption								√				√					**2**
Species		√													√		**2**
Forests						√									√		**2**
Innovation								√	√								**2**
Deaths			√								√						**2**
Recycling						√						√					**2**
Abuse			√													√	**2**
Count	**13**	**16**	**8**	**12**	**13**	**8**	**3**	**12**	**5**	**4**	**11**	**8**	**5**	**8**	**8**	**4**	

Source: developed by the author

Table 1 indicates that the top three keywords founded in the SDGs are (1) "Women" which is founded in 8 SDGs, (2) "Technology" which is founded in 7 SDGs, and (3) "Economic" which is founded in 7 SDGs. The focusing on such thematic area keywords may help decision makers to accelerate the speed to achieve the SDGs through adapting Nexus perspective. For example, the three keywords (Women, Economic, and Technology) are interlinked 12 SDGs. The adoption of a nexus approach, which integrates these three thematic areas, will allow to find effective and efficient solutions to tackle many problems and may achieve many targets of the SDGs. The Nexus perspective is essential for promoting the integration of SDGs across sectors and accelerating the decision-making process. Table 1 shows that, Goal 2 "End hunger" contains the highest number of thematic area keywords (16 keywords), while Goal 10 "Inequality" and Goal 16 "Peace" contains only 4 keywords for each of them.

Identifying SDGs Interlinkages

Once the matrix of Keywords-SDGs links is identified, it can be used to identify the SDGs Interlinkages matrix. Based on the concept of linguistic approach (Le Blanc, 2015; Miola et al., 2019), two goals are interlinked if they have in common a least one common thematic area keyword. Table 2 indicates the SDGs interlinkages matrix, through showing the number of keywords in common between each two goals. Table 2 can be directly concluded and calculated from Table 1. For example, Goal 1 and Goal 2 are interlinked since they are intersected and have in common 8 thematic area keywords. Goal 1 and Goal 3 are interlinked since they are intersected by 1 keyword, and so on.… Table 2 indicates that, the highest SDGs interlinked (intersected/interacted) and have in common keywords are:

1. Goal 1 "End poverty" and Goal 11 "Cities"; intersected by 9 keywords.
2. Goal 1 "End poverty" and Goal 2 "End Hunger"; intersected by 8 keywords.
3. Goal 4 "Education" and Goal 5 "Gender Equality"; intersected by 7 keywords.
4. Goal 2 "End Hunger" and Goal 4 "Education"; intersected by 6 keywords.
5. Goal 1 "End poverty" and Goal 5 "Gender Equality"; intersected by 6 keywords.
6. Goal 2 "End Hunger" and Goal 11 "Cities"; intersected by 6 keywords.

Table 2: SDGs interlinkages Matrix, the number of keywords interlinked SDGs

		SDGs																#
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Goals**
SDGs	**1**		8	1	5	6	2	2	3	2	2	9	2	3	3	2	1	15
	2	8		2	6	5	4	3	4	3	1	6	3	3	4	3	1	15
	3	1	2		2	1	2		2		1	3	2	1	3	1	2	13
	4	5	6	2		7	3	1	5	2	1	3		3	1		2	13
	5	6	5	1	7		2	3	4	3	3	3	1	1	4	2	2	15
	6	2	4	2	3	2			1			3	2	1	2	3		11
	7	2	3		1	3				2					2	1		7
	8	3	4	2	5	4	1			3	2	3	3	3	2	1	1	14
	9	2	3		2	3		2	3		1	1			2			9
	10	2	1	1	1	3			2	1		2			2			9
	11	9	6	3	3	3	3		3	1	2		3	3	1	1	1	14
	12	2	3	2		1	2		3			3			1	1		9
	13	3	3	1	3	1	1		3			3						8
	14	3	4	3	1	4	2	2	2	2	2	1	1			3		13
	15	2	3	1		2	3	1	1			1	1		3		1	11
	16	1	1	2	2	2			1			1				1		8
Total Number of Linkages																		184

Source: developed by the author

Table 2 indicates that (1) there is no goal is separated from other SDGs – All the SDGs are interlinked with each other. (2) The percentage of interlinkages between SDGs is 76.6%; since if all SDGs are fully interlinkages, we should have (16*15=240) potential interlinkages, however, we have only 184 interlinkages. (3) The minimum number of interlinkages is 7, which indicates that, the least interlinked goal is connected with half of SDGs. Based on such methods, it could be easier for decision makers to use such results to set SDG priorities for implementation phase.

Table 3 provides an aggregate figure of the interlinkages among SDGs through thematic area keywords. Based on Table 2, Table 3 ranks the 16 goals according to the number of other SDGs to which they are interlinked. Goal 1 "End poverty", Goal 2 "End hunger", and Goal 5 "Gender

Equality" top Table 3 and they have interlinkages with all other SDGs. At the bottom of Table 3 is Goal 7 "Energy" and Goal 16 "Peace".

Table 3: The interlinkages between the SDGs through keywords: an aggregated figure

Rank	SDGs	#Goals
1	Goal 1. End poverty in all its forms everywhere	15
2	Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture	15
3	Goal 5. Achieve gender equality and empower all women and girls	15
4	Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	14
5	Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable	14
6	Goal 3. Ensure healthy lives and promote well-being for all at all ages	13
7	Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all	13
8	Goal 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development	13
9	Goal 6. Ensure availability and sustainable management of water and sanitation for all	11
10	Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	11
11	Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation	9
12	Goal 10. Reduce inequality within and among countries	9
13	Goal 12. Ensure sustainable consumption and production patterns	9
14	Goal 13. Take urgent action to combat climate change and its impacts3	8
15	Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	8
16	Goal 7. Ensure access to affordable, reliable, sustainable and modern energy for all	7

Source: developed by the author

4.2.3 SDGs Influence Matrix

SDGs influence matrix helps decision-makers to concentrate on the critical SDGs which have great influence on other SDGs and considering the role of all SDGs (Scharlemann et al., 2020). The influence and dependence approach is one of the methods which is related to Cross Impact Analysis (Glenn & Gordon, 2009; Weimer-Jehle, 2006). Cross impact analysis is a general approach for evaluating the interactions/intersections between the most important influential factors (Glenn & Gordon, 2009). The SDGs influence matrix is a qualitative tool that could be used to provide a quantifiable measure to the level of interlinkages between SDGs (Kumar, Ahmed, Singh, & Sinha, 2018). The influence matrix allowed all the SDGs intersections/interactions to be considered (Scharlemann et al., 2020). Table 2 which represents a qualitative measure for the influence of each pairwise intersection/interaction, where, the number of common keywords (intersected keywords) between each two DSGs is listed at each row/column interaction (cell).

Table 4 shows the SDGs Influence Matrix, which represents the relative weight of intersections/influences between each two goals and calculated directly from Table 1 and Table 2; by dividing the number of common keywords of each goal at a row by the total number of keywords in the column of this goal. For example, the relative influence of Goal 1 on Goal 2

is equal to 8 (the common keywords between Goal 1 on Goal 2; as shown in Table 2) divided by 16 (count of keywords of Goal 2; as shown in Table 1), while, the relative influence of Goal 2 on Goal 1 is equal to 8 (the common keywords between Goal 1 on Goal 2) divided by 13 (count of keywords of Goal 2). The influence scores at each row/column intersection, which ranges from 0 (no interaction/influence) to 1 (high interaction/influence), can be considered the absolute interaction/influence strength.

Table 4: SDGs Influence Matrix

SDGs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	INF.
1		0.50	0.13	0.42	0.46	0.25	0.67	0.25	0.40	0.50	0.82	0.25	0.60	0.38	0.25	0.25	6.11
2	0.62		0.25	0.50	0.38	0.50	1.00	0.33	0.60	0.25	0.55	0.38	0.60	0.50	0.38	0.25	7.08
3	0.08	0.13		0.17	0.08	0.25		0.17		0.25	0.27	0.25	0.20	0.38	0.13	0.50	2.83
4	0.38	0.38	0.25		0.54	0.38	0.33	0.42	0.40	0.25	0.27		0.60	0.13		0.50	4.82
5	0.46	0.31	0.13	0.58		0.25	1.00	0.33	0.60	0.75	0.27	0.13	0.20	0.50	0.25	0.50	6.26
6	0.15	0.25	0.25	0.25	0.15			0.08			0.27	0.25	0.20	0.25	0.38		2.49
7	0.15	0.19		0.08	0.23				0.40					0.25	0.13		1.43
8	0.23	0.25	0.25	0.42	0.31	0.13			0.60	0.50	0.27	0.38	0.60	0.25	0.13	0.25	4.55
9	0.15	0.19		0.17	0.23		0.67	0.25		0.25	0.09			0.25			2.25
10	0.15	0.06	0.13	0.08	0.23			0.17	0.20		0.18			0.25			1.45
11	0.69	0.38	0.38	0.25	0.23	0.38		0.25	0.20	0.50		0.38	0.60	0.13	0.13	0.25	4.72
12	0.15	0.19	0.25		0.08	0.25		0.25			0.27			0.13	0.13		1.69
13	0.23	0.19	0.13	0.25	0.08	0.13		0.25			0.27						1.52
14	0.23	0.25	0.38	0.08	0.31	0.25	0.67	0.17	0.40	0.50	0.09	0.13			0.38		3.82
15	0.15	0.19	0.13		0.15	0.38	0.33	0.08			0.09	0.13		0.38		0.25	2.25
16	0.08	0.06	0.25	0.17	0.15			0.08			0.09				0.13		1.01
DEP.	3.92	3.50	2.88	3.42	3.62	3.13	4.67	3.08	3.80	3.75	3.82	2.25	3.60	3.75	2.38	2.75	

Source: developed by the author

The influence matrix of SDG implies the total influence power of each goal (in rows) which indicate the influence of that goal to achieve other SDGs. On the other hand, the dependence power (in columns) indicates the total influence of other SDGs, which may help to accomplish this goal. Table 4 shows the total influence (INF.) –on the right-hand side – and dependence power(DEP.)– on the bottom of Table 4 – of each SDGs on other SDGs.

4.2.4 MICMAC Analysis

Figure 3 shows the total dependence and influence scores for each SDG, which help to differentiate between the SDGs using MICMAC (Impact Matrix Cross-Reference Multiplication Applied to a Classification) by classifying the influence and dependence power of SDGs into four categories (Arcade, Godet, Meunier, & Roubelat, 2003). MICMAC is a structural analysis method which helps to graphically show the influence and dependence power of SDGs (Cai & Xia, 2018; Kumar et al., 2018). Based on Table 4, Figure 3 shows that there are four clusters –MICMAC Analysis-, which are explained as follows:

1. Cluster 1 – Autonomous SDGs (Excluded SDGs): These SDGs have a weak influence and also have a weak dependence power. No significant influence over other SDGs or not influenced by other SDGs. These SDGs have few numbers of thematic area keyword in common with other SDGs. This cluster have 5 SDGs. Goal 12 “Consumption and Production”, Goal 15 “Terrestrial Ecosystems”, Goal 16 “Peace”, Goal 6 “Water” and Goal 3 “Health”.
2. Cluster 2 – Dependent SDGs (Output/Result SDGs): These SDGs have strong dependence power but weak influence power. These SDGs have the strongest dependence power and

they (directly or indirectly) influenced by other SDGs. If the Driving and Linkage SDGs well-addressed, it will provide a good chance to achieve Dependent SDGs. This cluster contains five SDGs. Goal 13 "Climate Change", Goal 9 "Industry & Infrastructure", Goal 10 "Inequality", Goal 7 "Energy" and Goal 14 "Oceans".

3. Cluster 3 – Linkage SDGs (Relay/Conflict SDGs): These SDGs have a strong influence and dependence power. They are very sensitive and unstable, since any action on these SDGs will trigger an effect on other SDGs – Dependent SDGs – and also have feedback effect on themselves. This means they are very influential and highly vulnerable SDGs. We have 4 SDGs in this section. Goal 1 "End poverty", Goal 2 "End hunger", Goal 5 "Gender Equality", and Goal 11 "Cities".
4. Cluster 4 – Driving SDGs (Influential/Independent/Input SDGs): These SDGs have a strong influence power, but they have weak dependence. They are more capable of influencing other SDGs. These SDGs are very important to achieve other SDGs, because they have high influence on most of them and have little dependence on them; any change that occur in them will have a great effect on SDGs performance. This cluster has two SDGs: Goal 8 "Economic & Employment" and Goal 4 "Education".

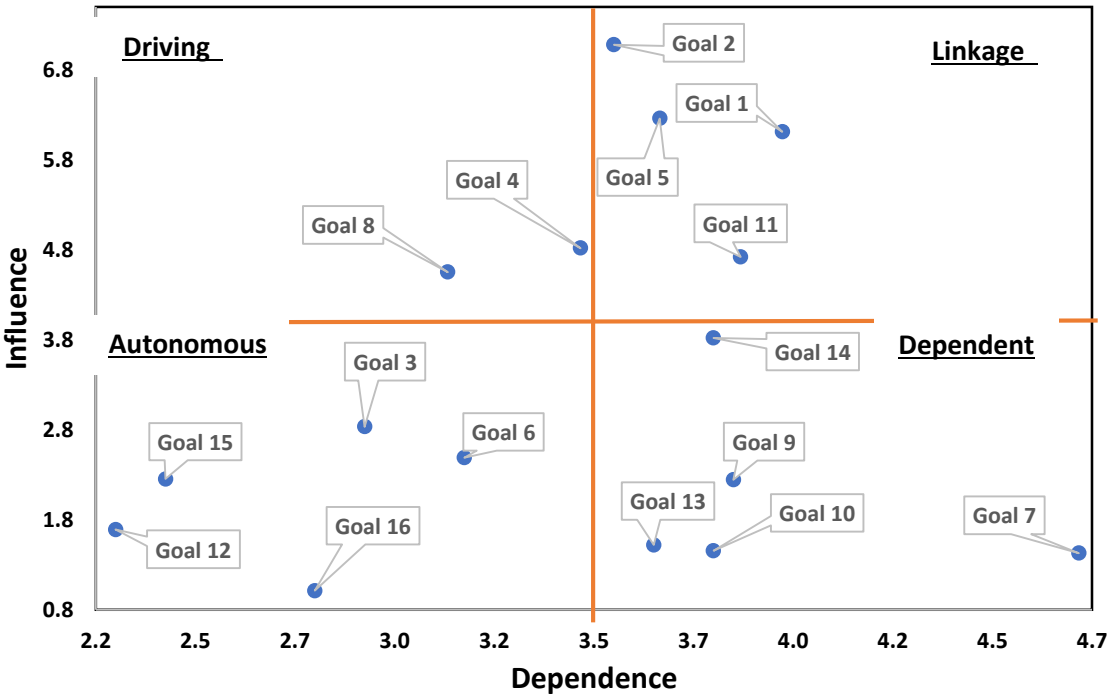

Figure 3: MICMAC Analysis

5. Conclusion

For developed and developing countries, there is a strong need for the decision-makers and policymakers to investigate and study the interlinkages of SDGs. Meanwhile, the launching of SDGs, and the Open Working Group emphasized on the need to understand the SDGs interlinkages; however, limited researches have been published to identify and quantify them. The importance of identifying the SDGs interlinkages is that prior to the implementation of a strategic plan to tackle and achieve the SDGs, it is necessary for the decision-makers to identify interaction, intersection and/or interlinkages between the SDGs and how they influence and

dependence on each other (Kumar et al., 2018). Also, these researches have used traditionally and manual methods.

In this paper, we proposed a text mining approach to identify the interlinkages of SDGs. The main thematic areas covered by the SDGs have been determined for each goal using, then the SDGs interlinkages have been concluded. The paper focuses on extracting the thematic area keywords using TF-IDF text mining method. Two goals are interlinked if they have in common at least one thematic area. The interlinkages presented here is limited to the explicit meaning of SDGs wording and based on the linguist approach. Also, The MICMAC analysis has been applied for further investigation for the SDGs interlinkages and to highlight the interrelationships between SDGs, since the planning to achieve one or more SDG should take into consideration other SDGs. The MICMAC analysis provides the policymakers with a working tool during their decision-making process. The influence power and dependence power calculated based on SDGs interlinkage matrix through keywords. As a result, the SDGs have been categorized into four clusters. Therefore, in light of the identification of the interlinkages and the MICMAC analysis, the policymakers should take into their account launch various managerial and strategic policies. Thus, the author suggests the following:

As a starting point for a strategic plan for SDGs; Goal 4 "Quality Education" and/or Goal 8 "Economy Growth" should be considered. Since they have a strong driving power; therefore they have a strong influence power to drive the achievement of other SDGs, as figure 3 shows. Meanwhile, the launching of SDGs. Education has always been as an integral part of the sustainable development agenda. Education is a key to escape from poverty and inspires a motivation to strive for the cause of sustainable development and to save the planet for generations to come. (Kumar et al., 2018) reported the same result. While, Economic Growth triggers for higher levels of productivity and technological innovation and also encourages for entrepreneurship and job creation. Therefore, it becomes imperative that the policymakers should start to modify their plans and develop strategies that could align with SDGs interlinkages.

The policymakers should also discuss other important factors while implementing SDGs strategic plans, such as financial budget, and social awareness. For example, this is important to achieve goals such as SDG 4 and SDG8 to drive the achievement of other SDGs. The development of an effective framework to clarify how to SDGs can be integrated is highly required.

This paper could provide a great value to decision-makers to understand more about SDGs interlinkages. The approach adopted here is very straightforward but very useful to get more insights about the SDGs interlinkages. This paper carried out an analysis at the global level. Similar analysis could be carried out at the national level, since countries such as Egypt should have different priorities, which depending on its national circumstances, it is likely to place varying emphasis on particular targets and goals.

ACKNOWLEDGMENT

I would like to acknowledge Prophet Muhammad (PBUH[4]) who has taken all human being from the darkness to the lightness.

References

[4] https://link.springer.com/referenceworkentry/10.1007%2F978-1-4020-8265-8_201281

- Arcade, J., Godet, M., Meunier, F., & Roubelat, F. (2003). Structural Analysis with the MICMAC Method & Actor's Strategy with MACTOR Method. Futures Research Methodology. *Laboratory for Investigation in Prospective Strategy , Paris*.
- Bakhtin, P., Khabirova, E., Kuzminov, I., & Thurner, T. (2020). The future of food production–a text-mining approach. *Technology Analysis & Strategic Management, 32*(5), 516-528.
- Buenaño-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *EEE Access, 8*, 35318-35330.
- Cai, Y., & Xia, C. (2018). Interpretive structural analysis of interrelationships among the elements of characteristic agriculture development in Chinese rural poverty alleviation. *Sustainability science, 10*(3), 786.
- Coopman, A., Osborn, D., Ullah, F., Auckland, E., & Long, G. (2016). *Seeing the whole: implementing the SDGs in an integrated and coherent way.* Paper presented at the Stakeholder Forum. London. UK.
- Cutter, A., Osborn, D., Romano, J., & Ullah, F. (2015). *Sustainable development goals and integration: Achieving a better balance between the economic, social and environmental dimensions.* Paper presented at the Stakeholder Forum.
- Ding, Z., Li, Z., & Fan, C. (2018). Building energy savings: Analysis of research trends based on text mining. *Automation in Construction, 96*, 398-410.
- Elder, M. (2016). *Strengthening the Linkages Between Air Pollution and the Sustainable Development Goals*: Institute for Global Environmental Strategies.
- Fabbrizzi, S., Maggino, F., Marinelli, N., Menghini, S., Ricci, C., & Sacchelli, S. (2016). Sustainability and food: a text analysis of the scientific literature. *Agriculture and agricultural science procedia, 8*, 670-679.
- Galati, F., & Bigliardi, B. (2019). Industry 4.0: Emerging themes and future research avenues using a text mining approach. *Computers in Industry, 109*, 100-113.
- Glenn, J. C., & Gordon, T. J. (2009). Futures Research Methodology Version 3.0.
- Jones-Diette, J. S., Dean, R. S., Cobb, M., & Brennan, M. (2019). Validation of text-mining and content analysis techniques using data collected from veterinary practice management software systems in the UK. *Preventive veterinary medicine, 167*, 61-67.
- Keeble, B. R. J. M., & war. (1988). The Brundtland report:'Our common future'. *4*(1), 17-25.
- Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts and Practice* (Vol. 2nd Edition): Morgan Kaufmann.
- Kumar, P., Ahmed, F., Singh, R. K., & Sinha, P. (2018). Determination of hierarchical relationships among sustainable development goals using interpretive structural modeling. *Environment, Development Sustainability, 20*(5), 2119-2137.
- Kwartler, T. (2017). *Text mining in practice with R*: John Wiley & Sons.
- Le Blanc, D. (2015). Towards integration at last? The sustainable development goals as a network of targets. *Sustainable Development, 23*(3), 176-187.
- Le Blanc, D., Freire, C., & Vierros, M. (2017). Mapping the Linkages between Oceans and Other Sustainable Development Goals.
- Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*: Academic Press.
- Miola, A., Borchardt, S., Neher, F., & Buscaglia, D. (2019). *Interlinkages and policy coherence for the Sustainable Development Goals implementation*. Retrieved from

- Nunes, A. R., Lee, K., & O'Riordan, T. J. B. g. h. (2016). The importance of an integrating framework for achieving the Sustainable Development Goals: the example of health and well-being. *1*(3).
- OWG, O. W. G. (2014). *Open Working Group on Sustainable Development Goals ANNEX 1*. Retrieved from https://sustainabledevelopment.un.org/content/documents/3387Annex_interlinkages_1903.pdf
- Park, K., & Kremer, G. E. O. (2017). Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems. *Ecological indicators, 72*, 803-820.
- Rivera, S. J., Minsker, B. S., Work, D. B., & Roth, D. (2014). A text mining framework for advancing sustainability indicators. *Environmental modelling software, 62*, 128-138.
- Scharlemann, J. P., Brock, R. C., Balfour, N., Brown, C., Burgess, N. D., Guth, M. K., . . . Wicander, S. (2020). Towards understanding interactions between Sustainable Development Goals: the role of environment–human linkages. *Sustainability Science*.
- Sebestyén, V., Domokos, E., & Abonyi, J. (2020). Focal points for sustainable development strategies—Text mining-based comparative analysis of voluntary national reviews. Journal of environmental management, 263, 110414.
- Shin, S.-H., Kwon, O. K., Ruan, X., Chhetri, P., Lee, P. T.-W., & Shahparvari, S. (2018). Analyzing sustainability literature in maritime studies with text mining. *Sustainability, 10*(10), 3522.
- Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*: John Wiley & Sons.
- Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach: " O'Reilly Media, Inc.".
- Team, R. C. (2018). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/
- UN. (2018). The Sustainable Development Goals Report 2018. *United Nations*.
- Vladimirova, K., & Le Blanc, D. (2016). Exploring links between education and sustainable development goals through the lens of UN flagship reports. *Sustainable Development, 24*(4), 254-271.
- Weimer-Jehle, W. (2006). Cross-impact balances: A system-theoretical approach to cross-impact analysis. Technological Forecasting Social Change 73(4), 334-361.
- Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction, 99*, 238-248.
- Zhou, X., & Moinuddin, M. (2017). *Sustainable Development Goals Interlinkages and Network Analysis: A practical tool for SDG integration and policy coherence*: Institute for Global Environmental Strategies.