

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

UTILIZACIÓN DE TÉCNICAS DE ANÁLISIS DE TEXTOS PARA EVALUAR INICIATIVAS SOCIALES Y ECOLÓGICAS DE SOSTENIBILIDAD EN EMPRESAS GLOBALES

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERÍA EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

Galo Anibal Moreta Páez

DIRECTOR: MsC. Carlos Montenegro Armas

CO-DIRECTOR: PhD. Edison Loza Aguirre

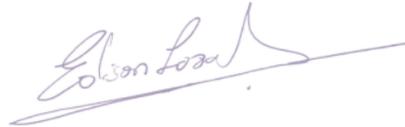
Quito, febrero de 2021

CERTIFICACIÓN

Certificamos que el presente trabajo fue desarrollado por Galo Anibal Moreta Páez, bajo nuestra supervisión.



MsC. Carlos Montenegro Armas
DIRECTOR DE PROYECTO



PhD. Edison Loza Aguirre
CO-DIRECTOR DE PROYECTO

DECLARACIÓN

Yo, Galo Anibal Moreta Páez, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



Galo Anibal Moreta Páez

DEDICATORIA

Dedico este trabajo de titulación a mis padres y hermano, por velar siempre por mí, estar en todas las etapas de mi vida y ahora verme alcanzar esta gran meta; a mis amigos que a lo largo de la carrera me supieron apoyar, y que han formado parte de este camino.

AGRADECIMIENTO

A mis padres por brindarme su apoyo, su cuidado a lo largo de los años, estoy muy agradecido por ello.

A mi hermano por estar en cada etapa de mi vida académica, y mi vida cotidiana con sus ocurrencias.

A mis tutores, Edison y Carlos por su apoyo y entrega en este proyecto, para que el mismo salga a flote en este tiempo que compartimos.

A mis amigos, los que fui conociendo en esta etapa y a los que ya venían conmigo desde mucho tiempo atrás, sinceramente gracias por sus granitos de arena.

ÍNDICE DE CONTENIDO

CERTIFICACIÓN.....	I
DECLARACIÓN.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
ÍNDICE DE CONTENIDO.....	V
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS.....	VIII
RESUMEN.....	IX
ABSTRACT.....	X
CAPÍTULO 1: INTRODUCCIÓN.....	11
1.1. Problemática.....	11
1.2. Objetivos.....	12
1.2.1. Objetivo General.....	12
1.2.2. Objetivos Específicos.....	12
1.3. Desarrollo Sostenible.....	12
1.3.1. Dimensiones del Desarrollo Sostenible.....	13
1.3.2. La Triple Línea Base de la Sostenibilidad.....	15
1.4. Trabajos relacionados.....	15
CAPÍTULO 2: METODOLOGÍA.....	18
2.1. Metodología de minería de textos en 4 etapas.....	18
2.1.1. Recolección de datos.....	19
2.1.2. Preprocesamiento de datos.....	21
2.1.3. Análisis de datos.....	22
2.1.4. Visualización de datos.....	22
CAPÍTULO 3: RESULTADOS.....	23
3.1. Resultados del análisis.....	23
3.2. Comparativa.....	26
CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES.....	28

4.1 Conclusiones	28
4.2 Recomendaciones	28
REFERENCIAS BIBLIOGRÁFICAS.....	30

ÍNDICE DE FIGURAS

Figura 1. Modelo de sostenibilidad de Elkington.....	15
Figura 2. Resultados sobre el estudio del equilibrio de DS.....	17
Figura 3. Aproximación de método tradicional de minería de textos.....	18
Figura 4. Representación del preprocesamiento de datos	21
Figura 5. Representación TBL final de la investigación.....	25
Figura 6. Comparativas de resultados TBL: a) Elkington, b) corpus de artículos académicos, c) corpus de reportes de empresas reconocidas mundialmente.....	26

ÍNDICE DE TABLAS

TABLA I. EMPRESAS COTIZADAS EN DOW JONES VS AÑOS RETENIDOS PARA EL ESTUDIO.....	20
TABLA II. NÚMERO DE REPORTES EMPRESARIALES RECOPIRADOS.....	20
TABLA III. CLASIFICACIÓN DE TÓPICOS.....	23
TABLA IV. AGRUPACIÓN DE TÓPICOS.....	24
TABLA V. PROBABILIDADES TOTALES PARA CADA DIMENSIÓN.....	25

RESUMEN

Diversas organizaciones han visto la necesidad de promover la incorporación de iniciativas de Desarrollo Sostenible, destacando la importancia de buscar un equilibrio entre los aspectos económicos, ambientales y sociales en las actividades que ellas realizan; hoy en día no existen métodos que permitan evidenciar que esto se esté logrando. En relación a este problema, en el presente proyecto buscamos evidenciar cuán equilibrado es el discurso en torno al Desarrollo Sostenible que las empresas mundialmente reconocidas presentan en sus reportes anuales de sostenibilidad. Las empresas elegidas para el estudio son aquellas que se han mantenido cotizando en los últimos 15 años en el índice Dow Jones de la bolsa de valores de Nueva York. Estos reportes fueron estudiados bajo un enfoque basado en Análisis de Tópicos, un método de Análisis de Textos, y el método Montecarlo con la finalidad de lograr una representación de cada una de las dimensiones de Desarrollo Sostenible de los reportes. Los resultados muestran que, al menos en el discurso, las empresas estudiadas declaran un mejor equilibrio dimensional, en comparación con estudios previos, con una prevalencia aún marcada hacia aspectos ambientales.

PALABRAS CLAVE: desarrollo sostenible, LDA, análisis de textos, TBL, Montecarlo.

ABSTRACT

Several organizations have pointed the need of promoting Sustainable Development initiatives, highlighting the importance of seeking a balance between economic, environmental and social aspects in the activities they develop. Nowadays, there are no methods that allow us to measure if this is being achieved. In this project, we seek to show how balanced is the discourse concerning Sustainable Development that world-renowned companies present in their annual sustainability reports. The companies chosen for the study are those that have been listed in the Dow Jones index for the past 15 years. These reports were studied under an approach based on Topic Analysis, a Text Analysis method, and the Montecarlo method to achieve a representation of the Sustainable Development dimensions as they appear on the reports. The results show that, at least in the discourse, the studied companies exhibit a better dimensional balance, compared to previous studies, with a still marked prevalence towards environmental aspects.

KEYWORDS: sustainable development, LDA, text analysis, TBL, Montecarlo.

CAPÍTULO 1: INTRODUCCIÓN

Antes de los años 70's, la gran mayoría de empresarios centraban su atención en conseguir únicamente ganancias financieras, dejando de lado aspectos ambientales y sociales. No fue sino hasta la ocurrencia de grandes desastres medio ambientales y las preocupaciones que estas generaban, que organizaciones mundiales y académicos plantean la necesidad de un desarrollo diferente, uno más respetuoso con el ambiente, y que pueda ser implementado a todo nivel, micro y macro económico [1]. Es así, como la Organización de las Naciones Unidas y expertos en el tema proponen varios principios que resguarden los bienes y recursos de manera que puedan estar disponibles sosteniblemente. Esto dio lugar así al término de "Ecodesarrollo" que después sería sustituido por el de "Desarrollo Sostenible" (DS) en los años 80's, para referirse al tipo de desarrollo que satisfaga las necesidades actuales sin comprometer a las generaciones futuras [2]. A su vez, surge la interpretación de DS como la búsqueda de un equilibrio entre 3 dimensiones: económica, ambiental y social [2].

1.1. Problemática

A pesar de que el DS es un tema de actualidad para gobiernos, empresarios, académicos y público en general, todavía hay una falta de mecanismos aceptados que ayuden a evaluar las iniciativas multisectoriales emprendidas en este ámbito. El presente proyecto pretende estudiar si existe un equilibrio, o no de las empresas mundiales en las prácticas de DS en los últimos 15 años. Por esta razón, para evaluar cada empresa se utilizará técnicas y procedimientos que ya han sido probados en el contexto académico [3, 4].

Las técnicas y procedimientos que se han planteado utilizar en el presente estudio están basadas en una aproximación de minería de textos, las cuales permiten extraer información útil e importante, como patrones y tendencias, a partir de un corpus. Así se adaptará una metodología de minería ejecutada en pasos secuenciales y basada en cuatro etapas: recopilación de datos (documentos de texto), limpieza de datos (representación estructurada), clasificación de datos (patrones y relaciones en el texto) y resultado de datos (patrones visualizados) [5].

Dentro de la etapa de clasificación de datos se utilizará una técnica de análisis de tópicos y Latent Dirichlet Allocation (LDA). Así, el conjunto de tópicos resultantes será analizado para establecer relaciones entre las tres dimensiones de DS. Estas relaciones serán integradas en una representación basada en el modelo de triple balance de

sostenibilidad (triple bottom line, TBL), gracias a una aproximación que utiliza el método de Montecarlo, con el fin de representar las iniciativas sostenibles; habilitando así la posibilidad de compararlas con resultados de estudios académicos anteriores.

1.2. Objetivos

1.2.1. Objetivo General

Utilizar técnicas de análisis de textos y visualización de datos para evaluar iniciativas sociales y ecológicas de sostenibilidad en las principales empresas globales que coticen en la bolsa de valores de Nueva York.

1.2.2. Objetivos Específicos

- Recolectar reportes de desarrollo sostenible de los últimos quince años de las principales empresas globales que coticen en el índice Dow Jones de Nueva York.
- Realizar un análisis de tópicos para identificar las temáticas más representadas en los reportes de desarrollo sostenible recolectados y clasificarlos de acuerdo a las tres dimensiones del desarrollo sostenible.
- Representar los resultados del análisis de tópicos según la representación del triple balance de sostenibilidad de Elkington de desarrollo sostenible utilizando el método de Montecarlo.
- Comparar los resultados obtenidos con estudios relacionados.

1.3. Desarrollo Sostenible

La noción de DS surge a partir de sustituir al término “Ecodesarrollo” en los años 80s como un concepto más completo, en el cual se incorpora temas sociales a las preocupaciones ecológicas y económicas que se generaban en la época [2]. Así, en el trabajo “Estrategia mundial de conservación” de la Unión Internacional para la Conservación de la Naturaleza y de los Recursos Naturales publicada en 1980 [6], el DS fue definido como el “tipo de desarrollo que proporciona mejoras reales en la calidad de vida humana y al mismo tiempo mantiene la vitalidad y diversidad de la Tierra. Hasta el día de hoy, esta noción parece utópica y, sin embargo, parece factible de operacionalizar al haber cada vez más personas que están convencidas de que esta es la única opción racional de desarrollo” [6].

En relación con la definición anterior, en el año 1981 el DS fue definido por Ignacy Sachs, una de las referencias más notables en la materia, como un desarrollo alternativo “que es socialmente deseable, económicamente viable y ambientalmente prudente” [7]. En

dicha definición, se estableció por primera vez la necesidad de atender las tres dimensiones de DS, social, económica y ambiental, con un principio de equilibrio entre las mismas.

En el informe de 1987 “Nuestro futuro común” [1], también conocido como informe Brundtland, se refuerza la relación indisoluble entre las tres dimensiones de DS. Dicho informe fue resultante del trabajo de la Comisión Mundial sobre Medio Ambiente y el Desarrollo creada por la Organización de las Naciones Unidas. El informe da a lugar una definición de DS que introduce además el principio de temporalidad a los conceptos ya existentes hasta esa fecha, planteando que “el desarrollo sostenible es un desarrollo que satisface las necesidades del presente sin comprometer la capacidad de las generaciones futuras para satisfacer las suyas”.

Finalmente, la Conferencia sobre el Medio Ambiente y Desarrollo de las Naciones Unidas celebrada en Río de Janeiro en el año de 1992 [8], tomó como referencia el informe Brundtland para dar lugar a la Declaración de Río sobre el Medio Ambiente y el Desarrollo, como un compromiso a nivel internacional hacia los principios de DS, que entraron en vigor en 1994 y han guiado los esfuerzos globales en la materia.

1.3.1. Dimensiones del Desarrollo Sostenible

Para Comeliau [9], el concepto de DS se lo puede manejar desde dos perspectivas:

- Las “ecocentradas” o “biocentradas” que tiene como objetivo esencial otorgar la protección de la vida en general,
- Y las “antropocéntricas” que busca el bienestar del ser humano.

Debido a estos enfoques y a la declaración de Río bajo el principio de que “los seres humanos están en el centro de las preocupaciones por el Desarrollo Sostenible, los mismos que tienen derecho a una vida productiva y sana en armonía con la naturaleza” (ONU, 1992), la definición de DS por parte de la Comisión Brundtland cae dentro de la perspectiva antropocéntrica.

Según Dion y Wolff [10], no se puede tomar al DS como una noción biocéntrica ya que impediría el desarrollo económico y la especie humana no sería más importante que otras especies a los ojos de la población. Para ellos, la perspectiva que dan al DS es antropocéntrica porque, según los autores citados, desarrollo económico y humano es lo mismo, y lo que se pretende es combinar estos desarrollos con una mejor protección del medio ambiente.

Por consiguiente, desde la perspectiva antropocéntrica, la protección del medio ambiente recae en garantizar que el mismo sea saludable y viable para futuras generaciones, dando lugar a la siguiente concepción de valor: sólo el ser humano tiene valor intrínseco, todo lo demás cobra valor porque el ser humano se lo atribuye en base a su capacidad de servir a la satisfacción de necesidades, intereses y deseos de este último.

A partir de la perspectiva antropocéntrica se puede identificar tres enfoques prácticos que surgen de la interpretación de las definiciones de DS, los cuales para Comeliau [9] son los siguientes:

- Enfoque económico, se da al hablar de la preservación del stock global ya sea natural o artificial, mediante la no explotación de aquellos recursos que no sean renovables.
- Enfoque ecológico, que reconoce la sustituibilidad del capital natural o artificial y que valida la explotación de los recursos a condición de su reposición con otro.
- Enfoque radicalista o sociocultural, que remarca la necesidad irrenunciable de considerar el rol de la satisfacción de las necesidades humanas, marcando una diferencia entre crecimiento y desarrollo.

A su vez, las nociones que se tiene acerca del DS revelan la necesidad de reconocer y tomar en cuenta la tensión entre sus tres dimensiones:

- Dimensión Económica: indica que el DS no puede ser ajeno a la necesidad de buscar mejoras o ganancias económicas. Hace énfasis en la diferencia entre desarrollo y crecimiento, donde el primero se refiere a un proceso de la transformación o evolución, mientras que el segundo es un simple incremento de capitales financieros. De la misma forma, vela por un desarrollo económico que satisfaga las necesidades de la población.
- Dimensión Ambiental: se centra en la defensa del medio ambiente, promulgando el respeto a la naturaleza mediante acciones que eviten la alteración significativa de los sistemas naturales debido a los actos del ser humano.
- Dimensión Social: se enfoca en las interacciones entre personas y a la vida en sociedad; haciendo énfasis en aspectos como la salud, educación, equidad, vivienda y seguridad. Promulga la satisfacción de las necesidades de los seres humanos como parte integral del sistema.

1.3.2. La Triple Línea Base de la Sostenibilidad

En el año de 1994, John Elkington acuñó el término de la Triple Línea de Base [11], o Triple Resultado (Triple Botom Line, TBL), como una representación visual y práctica de las interacciones entre las diferentes dimensiones del DS (Figura 1). En consideración microeconómica, la TBL puede ser utilizada como un marco de sostenibilidad que examine el impacto social, medioambiental y económico de las operaciones de una empresa. Por consiguiente, varias organizaciones han adoptado TBL para evaluar su desempeño en materia de DS como una perspectiva y mecanismo que ayude a crear mayor valor comercial [12].

En la Figura 1, la intersección de las tres dimensiones de DS (zona gris) representa el equilibrio deseable, o el 'verdadero DS'. Se considera así que mientras más grande sea el área de intersección, más equilibradas serán las iniciativas de DS multisectoriales [13].

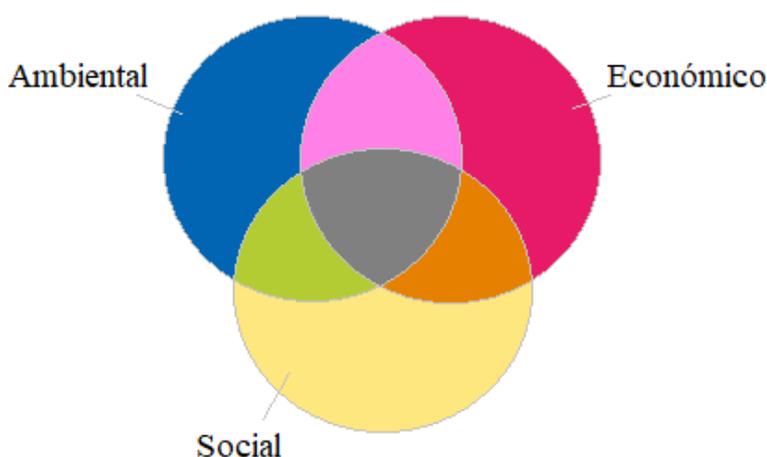


Figura 1. Modelo de sostenibilidad de Elkington.

Gracias a esto, se ha tomado TBL como mecanismo para medir cómo se manejan las empresas en cuestiones de DS. En ésta investigación, una representación TBL se conformará acorde a los resultados obtenidos en el análisis del corpus de reportes de DS publicados por las empresas.

1.4. Trabajos relacionados

El estudio de la implementación de iniciativas de DS se ha realizado en varios campos y a la vista sus diferentes interpretaciones. Por ejemplo, en [14], se destaca la búsqueda de las primeras interpretaciones de DS para resaltar la importancia de este como un principio moral y proponerlo como mecanismo de ayuda a la toma de decisiones,

permitiendo así establecer objetivos, actividades y reglas. A su vez, en [15], los autores examinan lo complejo que es implementar el DS en conjunto con iniciativas de innovación organizacional. De esta forma, los autores estudiaron la relación de los componentes de DS con los procesos internos de una empresa; con el objeto de identificar qué factores limitan una innovación respetuosa del DS a través de un modelo que ellos proponen.

No existen muchos trabajos en la literatura que hayan abordado el estudio del equilibrio dimensional en las iniciativas de DS como tal. Más reducido aún, es el grupo de contribuciones que lo hayan hecho a partir del análisis de textos. Aquellos que han sido encontrados, y que hayan empleado esta aproximación, se describen a continuación.

En [3], los autores se centraron en estudiar el equilibrio entre las tres dimensiones de DS en los artículos científicos publicados en las mejores revistas referentes a la gestión de cadena de suministro y operaciones, proponiendo una herramienta basada en la metodología de Montecarlo para realizar una representación gráfica basada en TBL (Figura 2 a). Los autores trabajaron con un corpus de literatura científica recolectada entre 2003 y 2018. Los artículos recogidos fueron codificados manualmente según su enfoque hacia las tres dimensiones de DS. Los resultados obtenidos revelaron que, en la literatura académica, no se encuentran balanceadas las tres dimensiones. Además, se muestra que la mayoría de las contribuciones se orientan a aspectos que intersecaban temáticas ambientales y económicas, dejando de lado el aspecto social.

En [4], los autores continúan con el estudio del equilibrio de las tres dimensiones de DS sobre el mismo corpus analizado en [3], pero probando la implementación de técnicas de análisis de tópicos basados en modelos que usan Latent Dirichlet Allocation (LDA) para sustituir la codificación manual usada en su primer estudio. LDA permitió así automatizar y acelerar el análisis del corpus, llegando a un resultado muy similar al obtenido en el estudio que empleó una codificación manual. Los resultados derivados con la técnica automatizada (Figura 2 b) son así comparables a los resultados obtenidos mediante la técnica manual, con una ganancia en tiempo muy considerable.

Finalmente, en [16] y [17], los autores ponen en conocimiento los resultados de la implementación de redes neuronales competitivas con el fin de acelerar y aumentar la precisión del análisis sobre el corpus. De esta manera, se sintetizaron métodos manuales y semi-manuales por medio de estas redes, lo que permitió detectar regularidades y correlaciones en la entrada, y adoptar futuras respuestas a esas entradas en consecuencia. Los resultados (Figura 2 c y d) son coincidentes con los

estudios previos; sin embargo, frente al modelo LDA, se reporta una leve ganancia en el tiempo dedicado al análisis.

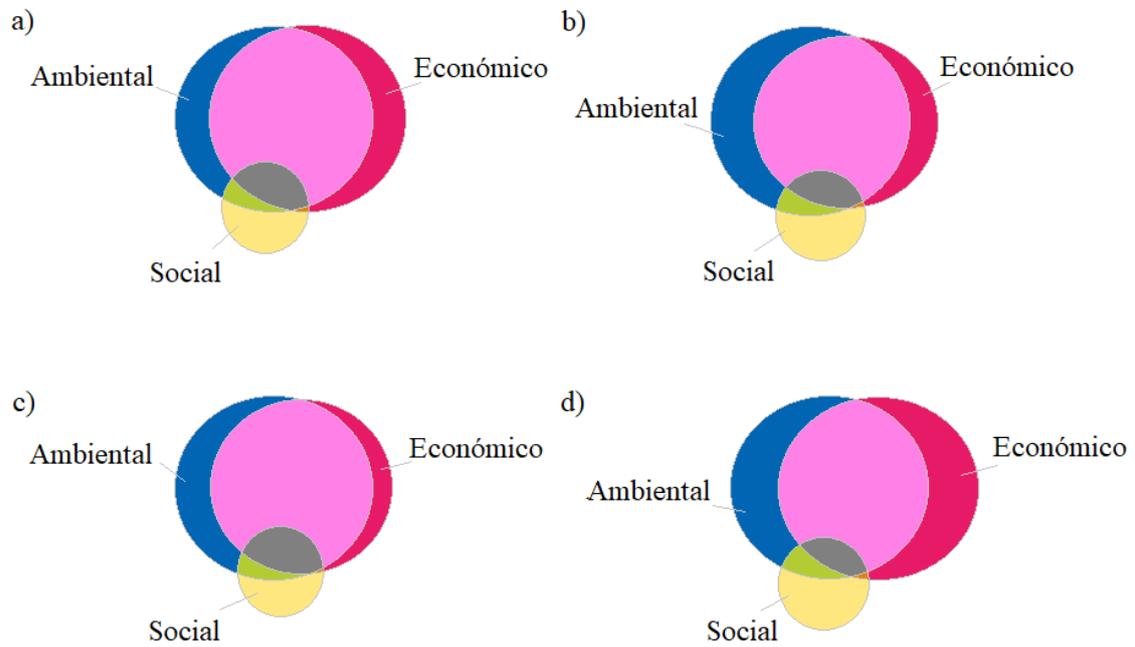


Figura 2. Resultados sobre el estudio del equilibrio de DS.

CAPÍTULO 2: METODOLOGÍA

2.1. Metodología de minería de textos en 4 etapas

Con el objetivo de estudiar si existe un equilibrio entre las tres dimensiones de DS en las prácticas de las empresas más reconocidas a nivel mundial, la presente investigación se enfoca en analizar los reportes empresariales relacionados con esta temática publicados anualmente por cada empresa. Para ello, se utilizará como metodología una aproximación de minería de textos basada en 4 etapas, lo que ayudará a sacar información útil e importante a través de un proceso sistematizado [5]. Este enfoque nos ayuda a cubrir pasos secuenciales como se puede apreciar en la Figura 3:

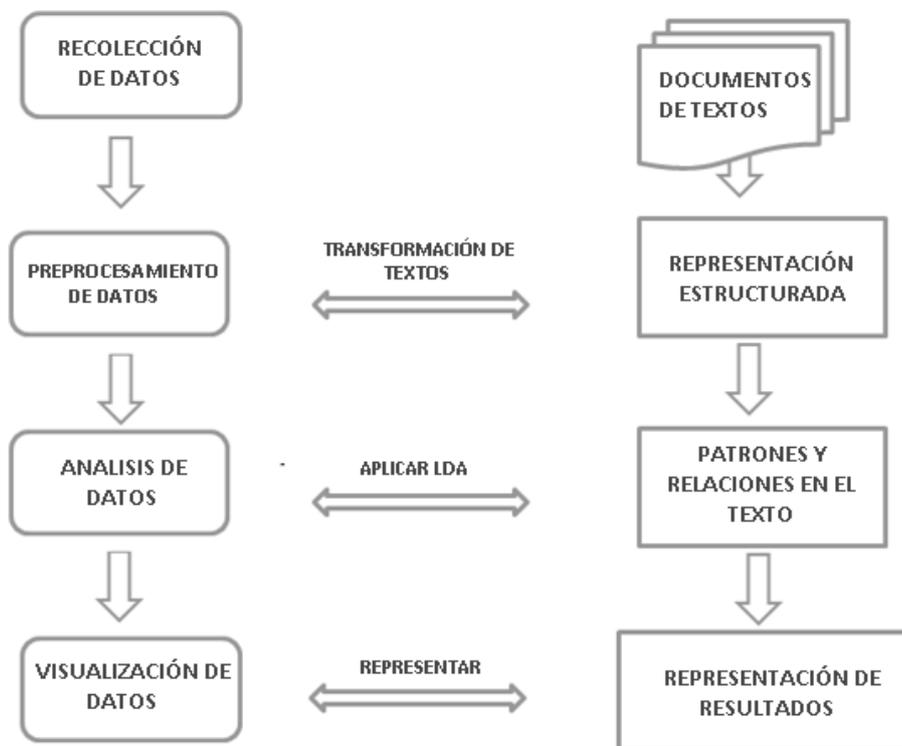


Figura 3. Aproximación de método tradicional de minería de textos.

En base a la metodología adoptada, las actividades y métodos implementados en cada etapa se describen a continuación.

2.1.1. Recolección de datos

A continuación, se lista los pasos que se realizó dentro de esta etapa:

Paso 1.- Identificar las empresas: Se realizó la búsqueda de las empresas que estén cotizando en el índice Dow Jones en el año 2019, como punto de partida.

Paso 2.- Elegir criterios de inclusión y exclusión: Para la recolección de reportes se eligió dos criterios de exclusión: el idioma, y si el reporte trataba acerca del desarrollo sostenible, para mantener un rango de publicaciones se tomó desde el año 2005 hasta el año 2019. Para el idioma se eligió inglés.

Paso 3.- Elegir términos de búsqueda para la recolección de reportes: Se utilizó términos relacionados con DS, como: Corporate Social Responsibility (CSR), Sustainability Report, Social Responsibility, Corporate Citizenship Report (CCR) y Environment Report.

Paso 4.- Relacionar los términos de búsqueda con las empresas: Para realizar la investigación, a más de utilizar estos términos, se añade el nombre de la empresa y el año de interés para encontrar los respectivos reportes de DS.

Paso 5.- Graficar tablas: Se elaboró dos tablas de resultados, la Tabla I muestra que existe un total de 30 empresas que han estado listadas en el índice Dow Jones, de las cuales 20 se han mantenido a lo largo de los últimos 15 años y la Tabla II refleja que en total se recolectaron 187 artículos provenientes de las 20 empresas cotizadas en Dow Jones retenidas para el estudio.

TABLA I. EMPRESAS COTIZADAS EN DOW JONES VS AÑOS RETENIDOS PARA EL ESTUDIO

Empresas	Años														
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
3M	Si														
American Express	Si														
Boeing	Si														
Caterpillar	Si														
Coca Cola	Si														
Exxon Mobil	Si														
Home Depot	Si														
IBM	Si														
Intel	Si														
Jhonson & Jhonson	Si														
JPMorgan Chase	Si														
McDonald's	Si														
Merck	Si														
Microsoft	Si														
Pfizer	Si														
Procter & Gamble	Si														
United Technologies	Si														
Verizon	Si														
Wall-Mart	Si														
Walt Disney	Si														
Apple	No	Si	Si	Si	Si	Si									
Chevron	No	No	No	Si											
Cisco	No	No	No	No	Si										
Dow Chemical	No	Si													
Goldman Sachs	No	Si													
Nike	No	Si													
Travelers	No	No	No	No	Si										
UnitedHealth	No	No	No	No	No	Si									
Visa	No	Si													
Walgreen	No	Si	Si	Si											

TABLA II. NÚMERO DE REPORTES EMPRESARIALES RECOPIRADOS

Empresas	N° Reportes
3M	8
American Express	4
Boeing	12
Caterpillar	9
Coca Cola	9
Exxon Mobil	9
Home Depot	5

IBM	16*
Intel	15
Johnson & Johnson	9
JPMorgan Chase	8
McDonald's	8
Merck	6
Microsoft	14
Pfizer	2
Procter & Gamble	15
United Technologies	3
Verizon	14
Wal-Mart	14
Walt Disney	7
Total, reportes a analizar	187

* En los reportes de IBM se encontró uno adicional que contempla el periodo 2007 - 2008.

2.1.2. Preprocesamiento de datos

Para el preprocesamiento de datos se siguieron los siguientes pasos:

Paso 1.- Transformar archivos PDF a textos planos: Se transcribió cada uno de los 187 reportes de manera manual a textos planos.

Paso 2.- Limpieza de caracteres: Se eliminó caracteres especiales de forma manual en cada uno de los textos planos con la finalidad de evitar errores en su posterior análisis. (Figura 4)

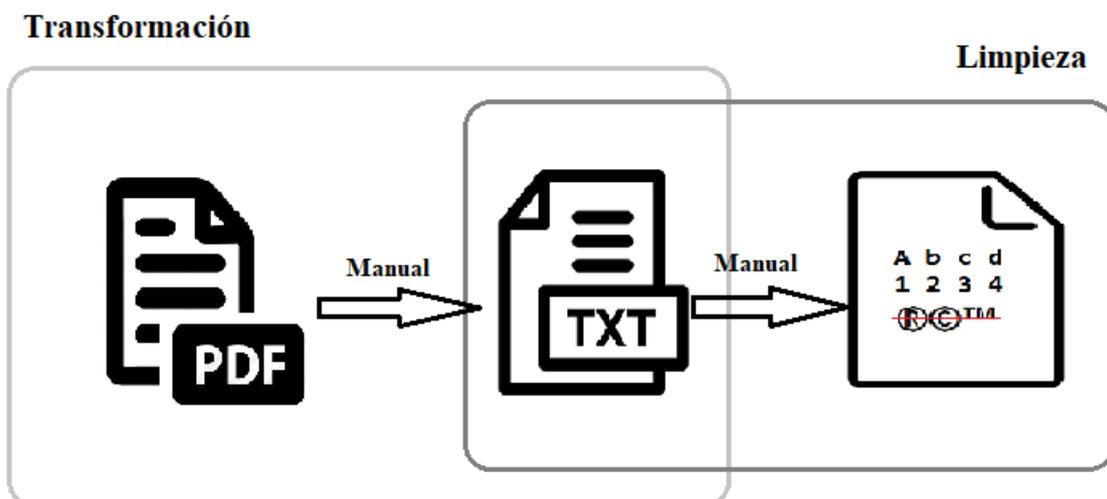


Figura 4. Representación del preprocesamiento de datos.

2.1.3. Análisis de datos

Codificación automatizada mediante Latent Dirichlet Asignation

Latent Dirichlet Asignation (LDA) es un modelo sin supervisión probabilístico el cual permite generar tópicos asumiendo que cada palabra en un documento es generada a partir de un tema al azar del corpus. La probabilidad de una secuencia de palabras no se ve afectada por el orden en el que aparecen [18].

A continuación, se describe los pasos para el análisis de datos:

Paso 1.- Tratamiento del corpus: Se ingresó el corpus preprocesado en la herramienta [4] para su estudio, que consiste en el procedimiento de análisis de tópicos, el cual permitió extraer automáticamente el significado del corpus mediante la identificación de temas recurrentes, que a su vez utiliza LDA, facilitando la identificación de tópicos constitutivos del corpus con sus respectivas probabilidades.

Paso 2.- Codificación del corpus analizado: Se codificó manualmente los resultados obtenidos por medio de un experto en estudios relacionados con DS, a fin de identificar la orientación de cada tópico en función de cada una de las dimensiones de DS, así como la de sus intersecciones.

2.1.4. Visualización de datos

Montecarlo

La aproximación de Montecarlo deriva de métodos matemáticos que se utilizaron alrededor de los años 40 para el desarrollo de armas nucleares. Técnicamente es un proceso estadístico numérico utilizado para resolver problemas matemáticos complejos a través de variables aleatorias [19].

Para la visualización de datos se realizó lo siguientes pasos:

Paso 1.- Tratamiento de resultados: Se identificó las probabilidades totales de cada categoría, así como en sus interacciones (categorías que correspondían a más de una dimensión) para mostrar las áreas de cada sección en una representación TBL.

Paso 2.- Representación de resultados: Con las probabilidades totales la siguiente herramienta [3] basada en el método de Montecarlo generó los círculos que representan cada dimensión de DS. modificando su posición y evaluando las áreas que se forman mediante un mecanismo de bombardeo de puntos, a fin de obtener una representación que, con el menor error posible, se aproxime a las probabilidades identificadas.

CAPÍTULO 3: RESULTADOS

3.1. Resultados del análisis

Como se evidenció en la Tabla II se recolectaron 187 reportes y a partir de estos se pasó a la clasificación y codificación dando lugar a los resultados de los tópicos identificados en el análisis, presentados en la Tabla III. En esta tabla se detalla la dimensión de DS, o dimensiones, codificadas por el experto en función de las palabras que constituyen el tópico, y que se listan también en la misma. Por otro lado, también se muestra la probabilidad de utilización del tópico en caso de ser buscado dentro de un documento del corpus. En este caso, esta probabilidad fue usada como indicativo de la participación del tópico dentro del conjunto de documentos.

TABLA III. CLASIFICACIÓN DE TÓPICOS

Tópico	Tópico 16	Tópico 10	Tópico 8	Tópico 20	Tópico 17
Dimensión	Económica	Econ. + Amb. + Soc.	Ambiental	Econ. + Amb. + Soc.	Econ. + Amb.
Probabilidad	0,3792	0,144337	0,045806	0,038137	0,032781
Palabras	business	employees	environment	employees	product
	global	products	environmental	education	energy
	million	management	emissions	corporate	water
	company	water	energy	environmental	environmental
	development	human	fuel	water	corporate
	year	environmental	waste	manufacturing	management
Tópico	Tópico 3	Tópico 22	Tópico 13	Tópico 4	Tópico 5
Dimensión	Social	Social + Ambiental	Econ. + Amb. + Soc.	Ambiental	Econ. + Amb. + Soc.
Probabilidad	0,03151	0,03043	0,02832	0,0237	0,02321
Palabras	rights	verizon	water	water	gas
	software	employees	waste	energy	oil
	human	customers	company	gas	safety
	online	energy	products	safety	training
	people	wireless	employees	power	economic
	privacy	services	children	emissions	management
Tópico	Tópico 2	Tópico 9	Tópico 24	Tópico 18	Tópico 25
Dimensión	Social	Social + Ambiental	Social + Ambiental	Social	Económica
Probabilidad	0,02296	0,02247	0,02216	0,02082	0,01968
Palabras	health	waste	water	johnson	capital
	merck	associates	bottling	health	business
	employees	food	partners	care	firm
	safety	stores	packaging	patients	financial
	group	foundation	local	healthcare	economic
	sites	sustainability	rights	healthy	investment
Tópico	Tópico 23	Tópico 11	Tópico 12	Tópico 21	Tópico 7
Dimensión	Econ. + Amb. + Soc.	Social + Ambiental	Social + Ambiental	Social + Ambiental	Econ. + Amb.
Probabilidad	0,01733	0,01711	0,01593	0,01429	0,01341
Palabras	mcdonald	disney	home	water	sustainability
	packaging	conservation	associates	employees	product
	sustainable	citizenship	oil	women	environmental
	food	families	factory	packaging	company
	markets	experiences	suppliers	oil	manufacturing
	children	nature	factories	consumers	customers
Tópico	Tópico 6	Tópico 14	Tópico 15	Tópico 1	Tópico 19
Dimensión	Económica + Social	Econ. + Amb.	Económica	Social	Económica + Social
Probabilidad	0,01259	0,01011	0,00704	0,00591	0,00076
Palabras	american	packaging	million	ibm	ranch
	card	stores	tax	employees	harvey
	employees	waste	approximately	clients	dating
	customers	products	cash	collaboration	emphasized
	service	energy	commercial	healthcare	innovations
	customer	factories	sales	responsibility	bought

De los grupos de palabras mostrados en la Tabla III, se puede observar que tan solo cinco de los veinticinco involucran palabras relacionadas con las tres dimensiones, once involucran dos de las tres dimensiones, y los nueve restantes se relacionan a solo una de las tres dimensiones. No obstante, el análisis que nos interesa se refiere principalmente a la probabilidad asociada a cada tópico. Así, en la Tabla IV se presenta la agrupación de los tópicos de acuerdo con la dimensión a la que se refieren, permitiendo el cálculo de la probabilidad acumulada para cada dimensión o intersección.

TABLA IV. AGRUPACIÓN DE TÓPICOS

Tópico	Probabilidad del Tópico	Dimensión DS	Probabilidad agregada
Tópico 8	0,0458	Solo ambiental	0,0695
Tópico 4	0,0237		
Tópico 16	0,3792	Solo económica	0,4059
Tópico 25	0,0197		
Tópico 15	0,007		
Tópico 3	0,0315	Solo social	0,0812
Tópico 2	0,023		
Tópico 18	0,0208		
Tópico 1	0,0059		
Tópico 17	0,0328	Ambiental	0,0563
Tópico 7	0,0134	+	
Tópico 14	0,0101	Económica	
Tópico 22	0,0304	Ambiental + Social	0,1224
Tópico 9	0,0225		
Tópico 24	0,0222		
Tópico 11	0,0171		
Tópico 12	0,0159		
Tópico 21	0,0143		
Tópico 6	0,0126	Económica +	0,0134
Tópico 19	0,0008	Social	
Tópico 10	0,1443	Económica	0,2513
Tópico 20	0,0381	+	
Tópico 13	0,0283	Ambiental	
Tópico 5	0,0232	+	
Tópico 23	0,0173	Social	

La Tabla IV nos permite construir la Tabla V, la cual contendrá el total de cada dimensión de DS (obtenida de sumar todas las probabilidades en las cuales aparece). Las probabilidades de cada componente mostradas en la Tabla V permitirán alimentar la herramienta de visualización basada en el método de Montecarlo para obtener la representación TBL del corpus, es decir su análisis final (Figura 5).

TABLA V. PROBABILIDADES TOTALES PARA CADA DIMENSIÓN

Ítem	Probabilidad
Total de componentes económicos	0,7269
Total de componentes ambientales	0,4995
Total de componentes sociales	0,4683
Total de componentes económicos y ambientales	0,3076
Total de componentes sociales y ambientales	0,3737
Total de componentes económicos y sociales	0,2647
Total de componentes compartidos en las tres dimensiones	0,2513

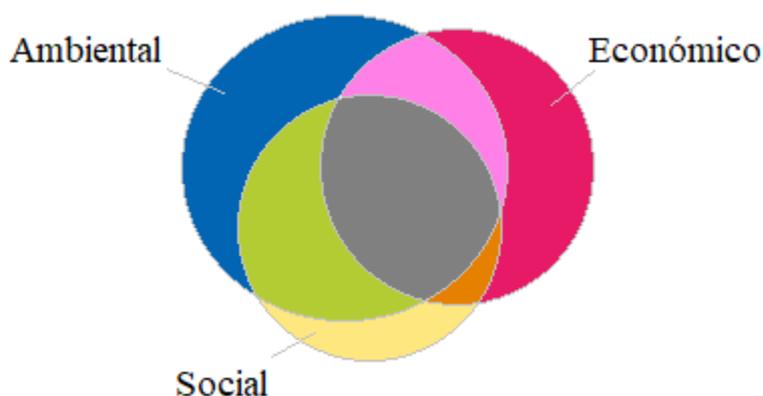


Figura 5. Representación TBL final de la investigación.

La Figura 5 muestra entonces una representación TBL en la cual se puede apreciar un cierto equilibrio entre las dimensiones de DS, la cual se caracteriza por:

- 1) La presencia de un tamaño relativamente equilibrado de las áreas que representan cada una de las dimensiones de DS. Todavía es posible evidenciar, sin embargo, que la dimensión Ambiental es la más representada por sobre las otras dos dimensiones.
- 2) Es importante notar que existe una relación muy estrecha entre la dimensión Ambiental y las otras dos dimensiones, desde la perspectiva de que casi todos los temas sociales y casi la mitad de los económicos se intersecan con la dimensión ambiental.
- 3) La zona de intersección de las tres dimensiones corresponde a casi la cuarta parte de todas las áreas, lo que significa que, al menos en el discurso, la conjunción de las tres dimensiones es entendida y practicada por las empresas analizadas.

3.2. Comparativa

Los resultados obtenidos en el presente estudio pueden ser comparados con los resultados reportados en estudios previos que utilizaron una aproximación metodológica similar, en la cual prácticamente solo cambia el corpus estudiado. Se puede centrar la comparación en analizar la situación ‘deseable’ en términos de DS, marcada por la conjunción, a la vez, de las tres dimensiones (es decir, las mejores iniciativas deberán aportar una ganancia económica, respetando el medio ambiente y posibilitando mantener o mejorar la situación social de las partes involucradas). Esta zona corresponde a la zona de triple intersección en la representación TBL de Elkington (parte gris de la Figura 6 a). Así, en el campo de lo deseable, lo ideal radica cuando todos los círculos se intersecan y obtenemos una representación TBL con una notable zona gris.

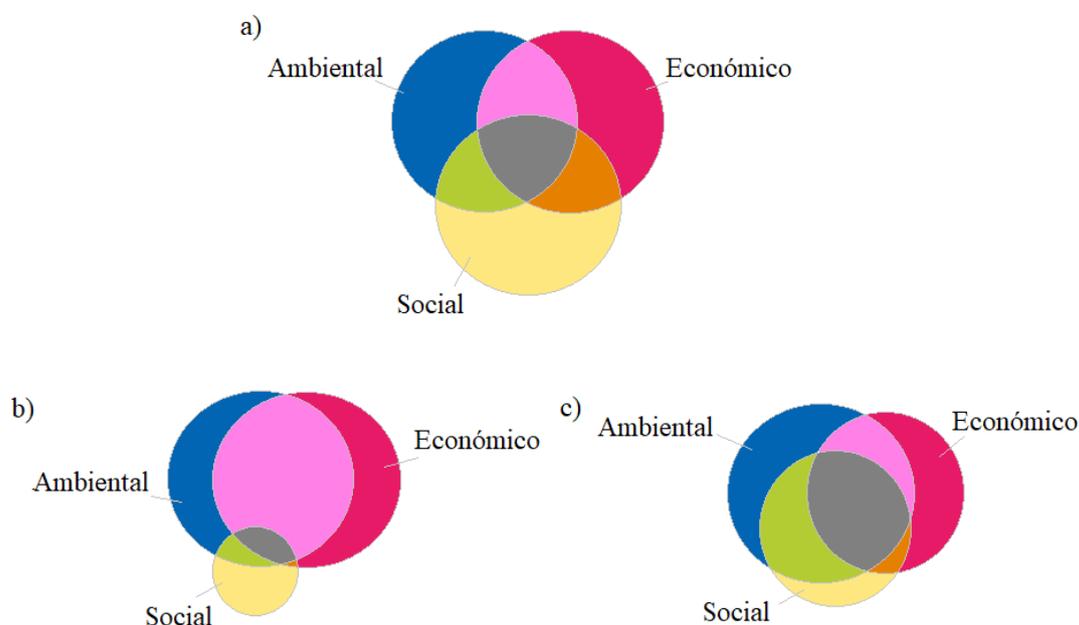


Figura 6. Comparativas de resultados TBL: a) Elkington, b) corpus de artículos académicos [15], c) corpus de reportes de empresas reconocidas mundialmente.

Si consideramos los resultados reportados en [3, 4, 16, 17] (Figura 6 b), en donde a partir del análisis de un corpus de artículos científicos, se observó una predominancia en aspectos Económicos y Ambientales, dejando muy de lado el aspecto social; los resultados del presente estudio (Figura 6 c) muestran diferencias notables. Se observa así que: el aspecto social se encuentra mucho más representado; el área de conjunción entre las tres dimensiones es mucho mayor; y el área económica deja de estar íntimamente ligada al área ambiental, transfiriendo un protagonismo más marcado a esta última dimensión.

Estos resultados dan lugar a dos reflexiones. Primero, que al menos en el discurso, las empresas globales tienden a incorporar esfuerzos que involucran a la vez las tres

dimensiones del DS, buscando en cierta manera un equilibrio y, por otro lado, se evidencia una desconexión entre la literatura académica y los reportes analizados. Lo que podría sugerir que el mundo académico no está necesariamente sintonizado con el mundo privado en cuanto a preocupaciones relacionadas con el DS, si aceptamos la presunción de que el discurso en los documentos refleja efectivamente las prácticas reales de estos actores.

CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- En la presente investigación, mediante un análisis realizado a un corpus de reportes de empresas reconocidas a nivel mundial, se estudió la existencia de un equilibrio entre las tres dimensiones de DS en las prácticas de esas firmas, o al menos en su discurso. El estudio se llevó a cabo sobre los reportes de DS de las 20 empresas que han cotizado sus acciones en el índice Dow Jones en los últimos 15 años. Los métodos de análisis utilizados incluyeron un análisis de tópicos basado en el algoritmo LDA, y la generación de una representación TBL apoyada por el método Montecarlo.
- De los resultados del análisis, se pudo demostrar la existencia de un equilibrio relativo entre las tres dimensiones en el corpus estudiado, con la dimensión ambiental siendo la más representada. Estos resultados evidencian un mejor resultado que aquellos reportados en estudios previos sobre un corpus de artículos científicos correspondientes al mismo periodo y a la misma temática, esto sugiere una diferencia de sintonía entre los centros de interés de la academia y los del sector privado. Futuros estudios pueden orientarse a explorar con más detalle esta divergencia y a incorporar el análisis de otros sectores, como el público.
- Es importante remarcar que el estudio realizado se refiere a un análisis del discurso expresado por las empresas estudiadas en sus reportes. Si bien, partimos de la premisa de que este discurso puede revelar las intenciones o las prácticas de los actores que lo producen, los resultados obtenidos pueden no necesariamente ser un reflejo de las prácticas reales. Así, futuros estudios pueden corroborar lo reportado mediante la triangulación de datos de otras fuentes como, por ejemplo, la inversión financiera destinada a cada dimensión del DS.

4.2 Recomendaciones

- Con relación a la metodología empleada, la misma se revela efectiva para evaluar las iniciativas de DS declaradas en los documentos. Sin embargo, posibles mejoras pueden aún ser introducidas para optimizar los métodos utilizados. Por un lado, la codificación manual de los tópicos en relación con la o las dimensiones a las cuales corresponden (que actualmente es realizada por un experto) podría en un futuro ser realizada a través de un modelo entrenado,

lo cual permitiría reducir tiempos de análisis y eliminar subjetividades en la codificación.

- Finalmente, falta aún trabajar en la introducción del concepto de temporalidad en los análisis. La aplicación de un análisis longitudinal puede permitir revelar evoluciones en las consideraciones y prácticas en torno al DS. Facilitando así posibles estudios sobre los impactos de influencias externas, sean estas normativas, coercitivas, o miméticas.

REFERENCIAS BIBLIOGRÁFICAS

- [1] ONU, "Our Common Future, Brundtland Report," 1987.
- [2] Naciones Unidas, "Informe de la conferencia de las Naciones Unidas sobre el medio humano," Estocolmo, 1972.
- [3] L. Edison, S. Marco, R. Henry, y M. Carlos, "Unveiling Unbalance on Sustainable Supply Chain Research: Did We Forget Something?," *Advances in Intelligent Systems and Computing*, Quito, Ecuador, 2018.
- [4] M. Carlos, L. Edison, y S. Marco, "Using probabilistic topic models to study orientation of sustainable supply chain research," *Advances in Intelligent Systems and Computing*, Quito, Ecuador, 2018.
- [5] M. Arispe, J. Capucac, F. Relucio y D. Maligat, "Teachers' sentiments to Bikol MTB-MLE: Using sentiment analysis and text mining techniques," *International Journal of Research Studies in Education*, vol. 8, nº 4, pp. 21-26, 2019.
- [6] UICN, PNUMA y WWF, "Estrategia mundial para la conservación," 1980.
- [7] I. Sachs, "Environment and Development Revisited," *Alternatives: Global, Local*, vol. 8, nº 3, pp. 369-378, 1982.
- [8] Naciones Unidas, "Declaración de Rio sobre el Medio Ambiente y el Desarrollo," Rio de Janeiro, 1992.
- [9] C. Comeliau, "Développement du développement durable ou blocages conceptuels?," *Tiers Monde*, vol. 35, nº 137, pp. 61-76, 1994.
- [10] Wolff D. Dion M., *Le développement durable: Théorie et applications au management.*, Paris: Dunod, 2008.
- [11] J. Elkington, "Partnerships from Cannibals with Forks: The Triple bottom line of 21st=Century Business," 1998.
- [12] T. Slaper y T. Hall, "The Triple Bottom Line: What Is It and How Does It Work?," *Indiana Business Review*. Spring 2011, Volume 86, No. 1, 2011.
- [13] C. Carter, P. Liane Easton, "Sustainable supply chain management: evolution and future directions," *International Journal of Physical Distribution & Logistics Management*, nº 1, vol 41.

- [14] A. Ross, "Modern Interpretations of Sustainable Development," *Journal of Law and Society*, vol 36, pp. 33-54, 2009.
- [15] A. Ponikarova, M. Zotov y A. Salin "Management model of sustainable development of petrochemical industry enterprises," *IOP Publishing Ltd*, vol. 1679, n° 5, 2020.
- [16] M. Carlos, L. Edison, y S. Marco, "Identifying the Orientations of Sustainable Supply Chain Research Using Data Mining Techniques: Contributions and New Developments," *Advances in Intelligent Systems and Computing*, Quito, Ecuador, 2019.
- [17] M. Carlos, L. Edison, S. Marco y Y. Eddy, "Core Topics Discovery in Sustainable Supply Chain Literature: An Automatic Approach," *Journal of Physics: Conference Series*, Quito, Ecuador, 2020.
- [18] D. Blei, y M. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, n° 4-5, pp. 993-1022, 2003.
- [19] K. Malvin, P. Whitlock, "Monte Carlo Methods," *Wiley*, 2008.