



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID



Graduado en Ingeniería Informática

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

TRABAJO DE FIN DE GRADO

Clasificador de textos de acuerdo con los Objetivos de
Desarrollo Sostenible usando Procesamiento de Lenguaje
Natural

Autor: Julen Pérez Álvarez

Director: Oscar Corcho García

Madrid, Junio de 2021

AGRADECIMIENTOS

A mi Aita, por inculcarme la belleza en la lógica y la razón, por sacrificarse tanto por darnos lo mejor, por estar siempre ahí y sacar tiempo para todo. Por ser el mejor padre. Te quiero y te admiro.

A mi Ama, por ser la madre más cariñosa y paciente de todas, por mirar siempre por el bien de los otros antes que el tuyo, por intentar siempre hacer feliz a todo el mundo. Por sacar tiempo para todo y estar siempre ahí. Te quiero y te admiro.

Nanito a ti también.

A Ele por sostenerme durante mis pequeños colapsos a lo largo de este proyecto, y durante los no tan pequeños. Por tus detalles inesperados y tu compañía. Tendríamos que habernos conocido antes. Me haces querer ser mi mejor versión.

A Carlos por su paciencia infinita con mis dudas sobre librAIry y a Oscar por sacar tiempo para y por hacer posible que todos tengamos un proyecto a nuestra medida.

A Enrique, Felipe, Iñigo y Juan, soy consciente que sin vosotros este camino de 4 años habría sido mucho más difícil y probablemente mas largo.

A Nacho, López, Almu y todo Amumus en general, por estar siempre ahí, tendré médicos de por vida.

Por último, a todas las personitas que forman parte de mi vida, que nunca van a leer esto y son demasiadas para nombrarlas, vosotras sabéis quienes sois.

Gracias.

Índice

AGRADECIMIENTOS.....	3
I Resumen.....	6
II Abstract	7
1 Introducción	8
1.1 Objetivos de desarrollo Sostenible	8
1.2 Planteamiento del Problema	9
1.3 Licencias:	10
2 Objetivos del Proyecto	11
2.1 Objetivo general.....	12
2.1.1 2.1.1 Seleccionar una fuente.....	12
2.1.2 2.1.2 Recopilación de los textos.....	12
2.1.3 2.1.3 Extracción de Datos de los textos creación del Modelo de Tópicos. Data Mining.	12
3 Impacto del trabajo.....	14
4 Estado del Arte	15
4.1 Trabajo de Fin de grado “Clasificación inicial de proyectos de una plataforma de crowdfunding de acuerdo con los Objetivos de Desarrollo Sostenible”.....	15
4.2 Organización de las naciones Unidas	15
4.2.1 UN SDG API.....	15
4.2.2 Naciones Unidas Biblioteca Digital.....	18
4.2.3 UN-iLibrary	19
4.3 Corresponsables.....	20
4.4 SDG PATHFINDER	21
5 Herramientas de Desarrollo	23
5.1 Python	23
5.2 librAlry.....	23
5.3 Docker	23
5.4 GitHub	23
5.5 Minería de textos	23
5.6 Web Scraping	24
5.7 Beautiful Soup	25
5.8 Web Drivers.....	25
5.9 Procesamiento de lenguaje natural	25
5.10 Labelled LDA.....	26
6 Desarrollo	28
6.1 Fase 1 Selección de una fuente	28
6.2 Fase 2 Recuperación de datos.....	29

6.3	Fase 1 y 2 bis	31
6.4	Fase 3: Extracción.....	34
6.5	Fase 4 Testeo del Modelo	40
6.5.1	Pruebas con textos extraídos de la ONU:.....	41
6.5.2	Pruebas con textos extraídos de la web de Corresponsables:.....	46
7	Conclusiones:	52
8	Bibliografía	54

I Resumen

Actualmente, la tecnología ha penetrado en todas las sociedades y se ha expandido a todos los niveles y en todos los ámbitos de nuestra vida a gran velocidad. Es un proceso imparable, pues cada vez van apareciendo nuevos avances, aplicaciones y elementos tecnológicos que incentivan a todo tipo de usuarios a hacer uso de ellos, facilitando la comunicación interpersonal, así como un acceso rápido y de forma instantánea a la información. En los últimos años la cantidad de datos e información producida y almacenada ha sufrido un crecimiento exponencial impulsado por tecnologías como la nube¹ y la facilidad de acceso a internet por la mayor parte de la sociedad, lo que supone un gran avance a la vez que surge un gran reto a la hora de categorizar toda esa información y filtrar la que nos interesa.

En el año 2015, la Organización de Naciones Unidas formulan los Objetivos de Desarrollo Sostenible (En adelante se utilizará su acrónimo ODS) para sustituir el plan de los Objetivos de Desarrollo del Milenio, que habían entrado en vigor desde el año 2000 y tenía fecha de finalización en ese mismo año. Los ODS conformaban un plan más ambicioso y actualizado que su predecesor. Este está formado por 17 objetivos de desarrollo, los cuales engloban unas 169 metas que dotan a los ODS de propósitos más concretos de alcanzar para el año 2030 con los que pretende reducir las desigualdades en el planeta y fomentar la preservación del planeta, apostando por la colaboración entre naciones.

Este proyecto surge con la idea de ayudar a identificar qué medidas se están tomando en cada uno de los 17 objetivos de cara a alcanzar las metas definidas en el plan de la ONU. El objetivo concreto de este proyecto es desarrollar un modelo que haga uso de tecnologías de análisis avanzado como es el Procesamiento del lenguaje Natural(NPL) y Machine Learning (ML) y que sea capaz de discernir si un texto trata sobre uno de los ODS y en caso afirmativo, que indique a qué ODS en concreto pertenece.

Desde el punto de vista técnico, el modelo de tópicos se entrenará con un corpus de textos marcados con cada uno de los 17 ODS, la mayor parte de estos textos han sido obtenidos de la página web de “corresponsables.com”², haciendo uso de Web Scraping y Web Drivers para automatizar la obtención de datos. Se ha escogido esta fuente por contener una gran cantidad de textos en castellano sobre el tema en cuestión, por la facilidad a la hora de acceder al contenido sin restricciones y por tener los textos ya clasificados de acuerdo con cada uno de los ODS.

El modelo de tópicos seguirá el modelo matemático de LDA, el cual es el modelo más utilizado desde el año 2012, con esto se pretende localizar la mezcla de palabras que se asocia con cada tópico, en este caso, cada tópico será un ODS.

II Abstract

Technology has become an integral part of our lives and our society; from the way we communicate to the way we do business. It is an unstoppable trend, as new emerging technologies arise, and users are keen to adopt this new development.

Equality, planet conservation and assuring prosperity are fundamental for our future as a society, it is of extreme importance to have a global strategy to tackle this problem and to achieve collaboration among nations.

On 2015, the United Nations came up with a 17 point plan to tackle this issue, they were called the “Sustainable Development Goals (SDGs)” and were meant to take over the previous plan “Millennium Development Goals” which was due that year. The SDGs were an update on the previous plan, making it more ambitious, learning from the mistakes of the last plan and building up on its strengths.

The SDG's are made up of 17 objectives, which are themselves divided in smaller goals, making up a total of 169 goals to achieve for 2030.

This project comes out of the necessity to identify what measures are been taken by not only the governments but also corporations to ensure the SDGs are a success. The goal of this project is to develop a model that identifies within which SDG a text fits, if any, this will be possible thanks to advances analysis technologies such as Natural language processing (NLP) and Machine Learning (ML).

The model Will be trained with 150 texts of each one of the SDG, the texts were obtained with Web Scraping and Web Drivers on the site “corresposables.com”. This source was chosen due to the vast number of texts in Spanish specific to the topic, because of the accessibility of this information and most importantly, because the texts were already tagged with a specific SDG.

The Topic Model Will be made using the Mathematical model of LDA, which is the most common one since the year 2012, and which will help determine which words are associated with which topic.

1 Introducción

1.1 Objetivos de desarrollo Sostenible

Los Objetivos de Desarrollo Sostenible nacen como parte importante de la Agenda de Desarrollo de 2030, en el año 2015 para renovar el recientemente finalizado plan de 15 años propuesto en el año 2000 (Objetivos de Desarrollo del Milenio). Esta nueva agenda de 2030 es ratificada por 193 estados miembro, teniendo un gran apoyo en la comunidad internacional, en él se detallan 17 Objetivos de desarrollo sostenible, los cuales cuentan con un total de 169 metas que dotan a la agenda de objetivos concretos, al mismo tiempo que establecen métricas para que se pueda evaluar el éxito y progreso de las naciones y corporaciones con respecto a cada objetivo.

Según los datos publicados por las naciones unidas, la inversión en el año 2019 fue de alrededor de 24.220 millones de dólares estadounidenses³, hay que tener en cuenta que este número sea probablemente mayor, ya que hasta el 31/12/2021 no se implantará el estándar por el cual han de hacerse públicos todos los gastos, así como el desglose en cada uno de los ODS. Sin embargo, esta cifra, debería darnos una idea de la magnitud de recursos destinados a la consecución de estos objetivos y la importancia de la Agenda de 2030 en nuestra sociedad actual.

Esta nueva agenda busca expresar el principio de responsabilidad común, donde todos los países participan y se enfrentan tanto a retos comunes como individuales para conseguir las diversas y amplias dimensiones del desarrollo sostenible resumidas en los ODS. A parte de su universalidad, la agenda 2030 proporciona una visión transformadora centrada en el planeta, en la dignidad de las personas y basada en los derechos humanos, además ofrecen una visión civilizadora entre los países y en el interior de estos, lo que conlleva la responsabilidad de los estados de promover la igualdad sin distinción de color, sexo, idioma, religión, opinión política, etc.

Por último, los ODS también constituyen una herramienta planificadora y de seguimiento de los países tanto a nivel nacional como local. Gracias a su visión a largo plazo constituirán un apoyo para cada país en su camino hacia un desarrollo sostenible e inclusivo, comprometido con el medio ambiente, a través de instrumentos de planificación, evaluación y presupuesto.

Los objetivos de Desarrollo Sostenible están listados a continuación:

1. Fin de la pobreza
2. Hambre cero
3. Salud y bienestar
4. Educación de calidad
5. Igualdad de género
6. Agua limpia y saneamiento
7. Energía asequible y no contaminante
8. Trabajo decente y crecimiento económico
9. Industria, innovación e infraestructura

10. Reducción de las desigualdades
11. Ciudades y comunidades sostenibles
12. Producción y consumo responsables
13. Acción por el clima
14. Vida submarina
15. Vida de ecosistemas terrestres
16. Paz, justicia e instituciones sólidas
17. Alianzas para lograr los objetivos

OBJETIVOS DE DESARROLLO SOSTENIBLE



Ilustración 1 Imagen con los 17 Objetivos de Desarrollo sostenible planteados por la ONU.

1.2 Planteamiento del Problema

La evolución de la tecnología en los últimos años ha experimentado un crecimiento exponencial, estos avances han servido para abordar problemas hasta la fecha inabordables; sin embargo, este progreso tecnológico, también ha traído consigo nuevos y apasionantes retos, con una base de personas conectadas a internet cada vez mayor, la penetración de internet rondaba el 59%⁴, y una generación de datos cada vez más amplia , siendo la cantidad de datos creados en todo el mundo en 2018 unos 33 zettabytes (un zettabyte equivale a 1.000 millones de terabytes), 16,5 veces más que solo hace nueve años y se espera que este crecimiento exponencial continúe esta tendencia alcista, una cantidad de datos, que simplemente escapa la capacidad de procesamiento por parte de los humanos.

A la vez que ha habido un incremento en el número de datos, también ha crecido el interés en analizar estos datos, y con ello una evolución en las tecnologías dedicadas a la recopilación y análisis de estos.

Al mismo tiempo que se ha incrementado el volumen de los datos y el acceso a ellos, se han ido formulando leyes que legislen el uso que puede hacer de ellos, de forma transparente y respetando la privacidad. La aparición de portales de datos abiertos son una buena herramienta para recopilar información sin inferir en incumplimiento de leyes de privacidad de datos ni violaciones a la propiedad intelectual.

Gracias a esta y otras herramientas hoy en día ya somos capaces de sacar conclusiones de grandes volúmenes de información, encontrar patrones que son a simple vista difícilmente detectables, y esto es exactamente lo que pretendemos en este proyecto, inferir características de cada uno de los ODS para poder distinguirlos entre ellos.

Muchos investigadores y profesionales del sector de la informática coinciden en la dificultad a la hora de gestionar el gran volumen de datos a nuestra disposición, a la vez que todos ellos entienden la relevancia que esto podría tener para sus líneas de trabajo, y las oportunidades que traería, pero para ello ha de mejorar el acceso a dichos datos. Es por eso por lo que el manejo de grandes volúmenes de datos es una de las líneas de investigación más importante en la actualidad, y es comúnmente conocida como Big Data.

En el caso concreto que nos compete en este proyecto, a pesar de la gran importancia que tienen los Objetivos de Desarrollo Sostenible y la gran cantidad de datos que estos producen, no hay una base de datos que permita filtrar los documentos por ODS, ni una manera sencilla de extraer una gran cantidad de temas categorizados con los ODS, de la misma forma, no hay un ninguna herramienta que permita clasificar un texto según su ODS, es justamente eso lo que se propone en este trabajo, por esto creemos que este proyecto puede ser de gran utilidad para nuestra sociedad actual.

1.3 Licencias:

Hay multitud de licencias de textos, en este proyecto nos vamos a centrar en las licencias digitales más importantes y como estas afectan a la hora de escoger fuentes de información.

Las licencias son instrumentos legales por los cuales se busca proteger los derechos de autor sobre las obras intelectuales⁵. Existen diversos tipos de licencia que otorgan una amplia variedad limitaciones en la circulación, distribución y comercialización de las obras.

Es relevante hablar del tipo de licencias porque hemos de escoger una fuente cuyas licencias se adecuen con nuestras necesidades, es decir, necesitamos una fuente que nos permita recopilar y modificar información, aunque no nos permita distribuirla y comercializarla. A continuación, pasaremos a listar algunos de los tipos de licencia digitales más relevantes:

- Copyright, licencia más restrictiva, su símbolo es una C en un círculo y viene acompañado de unas letras que indican que todos los derechos están reservados para el autor o dueño de la obra, esto quiere decir que todo uso, modificación o distribución del contenido queda amparado por el Copyright, y debe ser autorizado por el autor, muchas veces a cambio de una contribución monetaria. Ha de destacarse que es la licencia por defecto en caso de no indicarse una, aun que pueden usarse con limitaciones con fines educativos.



Ilustración 2 Imagen que muestra el sello de Copyright

- Copyleft, licencia opuesta en cuanto a restricciones que el copyright, es una licencia abierta que permite todo uso de la obra y versiones derivadas simplemente dando crédito al autor, sin autorización expresa de uso. Su logo es una C invertida (mirando a la izquierda) en un círculo.



Ilustración 3 Logo de Copyleft

- CreativeCommons: Esta licencia es un punto intermedio entre las dos anteriores, está pensada para permitir el uso de las obras, pero con ciertas condiciones, engloba un espectro amplio de restricciones, pero siempre bajo la idea de libre uso, pero con algunos derechos reservados para el autor. Su logo son dos C englobadas por un círculo.

TIPOS DE LICENCIAS CREATIVE COMMONS (CC)	
Reconocimiento (BY) BY Permite cualquier explotación de la obra, incluyendo una finalidad comercial, así como la creación y distribución de obras derivadas sin ninguna restricción.	Reconocimiento - NoComercial (BY-NC) BY NC Permite la generación de obras derivadas sin uso comercial de la obra original.
Reconocimiento - NoComercial - Compartirlugal (BY-NC-SA) BY NC SA No se permite uso comercial de la obra original ni de las posibles obras derivadas, cuya distribución debe hacerse con una licencia igual a la que regula la obra original.	Reconocimiento - NoComercial - SinObraDerivada (BY-NC-ND) BY NC ND No se permite un uso comercial de la obra original ni la generación de obras derivadas.
Reconocimiento - Compartirlugal (BY-SA) BY SA Se permite el uso comercial de la obra y de las posibles obras derivadas, cuya distribución debe hacerse con una licencia igual a la que regula la obra original.	Reconocimiento - SinObraDerivada (BY-ND) BY ND Se permite el uso comercial de la obra pero no la generación de obras derivadas.

Ilustración 4 Imagen que muestra los tipos de licencia Creative Commons, extraída del proyecto Cecarm⁶

Ha de tenerse en cuenta que las licencias, aunque relevantes, no son definitivas, según el aviso legal de Google sobre derechos de autor, hay determinados casos donde es posible utilizar "Obras" aunque estas estén sujetas a derecho de autor, ya tendríamos que indagar materia legal y en temas como la naturaleza del trabajo copiado, la intención y si el trabajo resultante de utilizar la obra con copyright puede o no llegar a sustituirla.

2 Objetivos del Proyecto

2.1 Objetivo general

Por todo lo expuesto en el apartado “Planteamiento del problema”, consideramos que resulta interesante crear una herramienta que permita identificar qué textos abordan temas relacionados con los ODS, esta herramienta podría ser utilizada de cara a ver realmente qué medidas se están tomando para lograr los objetivos propuestos, así como el avance real que se está produciendo en los diferentes países y corporaciones que se han adherido al plan de Desarrollo de 2030. Es por eso por lo que el objetivo de este proyecto es la creación de un clasificador de textos.

2.1.1 2.1.1 Seleccionar una fuente.

Para ello, el primer paso que tenemos que tomar es encontrar una fuente de información que nos proporcione una gran cantidad de textos etiquetados con los ODS, la elección de esta fuente resultará de vital importancia para el desarrollo del proyecto, es por esto por lo que se ha establecido que la fuente a utilizar cumpla los siguientes parámetros:

- Ha de ser una fuente abierta que permita el acceso y uso de sus datos.
- Ha de tener información de fácil acceso, a ser posible a través de una API o similar.
- Ha de tener una gran cantidad de textos para poder analizar.
- Ha de tener los textos catalogados dentro de las ODS.
- Ha de tratar temas de relevancia para los ODS.

2.1.2 2.1.2 Recopilación de los textos.

Los textos han de ser traídos desde su fuente para poder ser posteriormente procesados. La extracción de los textos puede realizarse de diversas formas, bien sea a través de peticiones API, descargas directas de archivos o incluso Web Scraping. Indagaremos más en este tema en la sección 4 “Herramientas de desarrollo”.

2.1.3 2.1.3 Extracción de Datos de los textos creación del Modelo de Tópicos. Data Mining.

Una vez obtenido un corpus de textos de los ODS, usaremos los textos para obtener patrones difícilmente identificables por los humanos, y a través de la tecnología de Data Mining, procesar esta gran cantidad de textos, y con ellos procederemos con el entrenamiento de los modelos de tópicos correspondientes, haciendo uso de técnicas de Machine learning y Labelled LDA, el cual describiremos en el apartado “estado del arte”.

Los subgrupos en los que clasificar los textos son una parte muy importante a la hora de generar un modelo basado en técnicas de procesamiento de lenguaje natural, sin embargo, para este proyecto se usarán los ODS como categorías, por lo que el trabajo previo en este campo no será necesario.

2.1.4 Testeo del modelo de tópicos y construcción de una interfaz WEB.

Finalmente procederemos a comprobar la eficacia del modelo de tópicos creado, a través de pruebas con textos que, estando clasificados en alguno de los ODS, no hayan formado parte de la muestra de entrenamiento, se analizarán los tokens obtenidos para comprobar que efectivamente guardan relación con los ODS específicos.

Para facilitar las pruebas y hacerlas accesibles para todo el mundo se planea desarrollar una interfaz web sencilla que permita, introduciendo un texto, devolver el ODS al que pertenece dicho texto.

3 Impacto del trabajo

Consideramos que este proyecto tiene una relación directa con los Objetivos de desarrollo sostenible al estar estrechamente ligado al tema de estudio. Estamos comprometidos con la idea de poder ayudar a la comunidad ofreciendo una herramienta que sea capaz de clasificar un texto según el objetivo de desarrollo sostenible al que pertenece.

Creemos que puede tener un impacto positivo para la sociedad, al permitir realizar en cuestión de segundos una tarea que de lo contrario requeriría de la lectura y comprensión del documento, esto permitiría reducir significativamente el tiempo empleado por individuos y organizaciones en realizar estudios y análisis sobre los ODS.

Hemos construido un buen corpus de entrenamiento sobre el cual puede seguir construyéndose y que es capaz de realizar una actualización auto-incremental, por lo que el modelo puede y debe ir creciendo en número de textos y previsiblemente aumentar la efectividad del clasificador.

Creemos que facilitaría la labor de evaluar el progreso y las medidas tomadas por los gobiernos y organizaciones en el desarrollo de los ODS y al haber realizado una interfaz web, creemos que cualquier usuario no experto podrá hacer uso de la herramienta de una manera sencilla.

En conclusión, creemos que el impacto que puede tener este proyecto es muy positivo, aun que se deba seguir construyendo sobre él, creemos que establece una buena base sobre la que desarrollar en futuros proyectos.

4 Estado del Arte

4.1 Trabajo de Fin de grado “Clasificación inicial de proyectos de una plataforma de crowdfunding de acuerdo con los Objetivos de Desarrollo Sostenible”⁷

El actual trabajo es un seguimiento del Trabajo de Fin de Grado realizado por el estudiante Alberto Sánchez Ruiz, en el año 2019. En dicho trabajo, se realizó un “Prototipo de clasificador de proyectos de la plataforma de crowdfunding Goteo.org de acuerdo con los Objetivos de Desarrollo Sostenible de las Naciones Unidas”. En este proyecto, el autor recopiló proyectos de micro financiación y los enmarcó dentro de los ODS, con esos textos siguió un proceso muy similar al que seguiremos para la tokenización de los textos.

En el TFG que estamos realizando construiremos sobre ese proyecto ya realizado , aprendiendo así del contenido útil hasta el momento y recopilando un corpus de textos de mayor tamaño, dado que esa fue la principal sugerencia de mejora por parte de su autor.

4.2 Organización de las naciones Unidas

Las Naciones Unidas es la Organización impulsora de los Objetivos de Desarrollo de 2030, cuenta con el apoyo de 193 estados miembros y emplea miles de personas. Como tal es una de las mayores fuentes de datos referentes a los ODS, incluyendo charlas, debates artículos y tratados, muchos de ellos en formato texto.

Durante el desarrollo de este proyecto, nos hemos encontrado con tres páginas las cuales parecen tener más relevancia como fuentes de información para generar el corpus, estas son:

4.2.1 UN SDG API⁸

Cuenta con una API desde la que se puede obtener información sobre los diferentes aspectos de los Objetivos de desarrollo sostenible, desde información referente a las distintas áreas de acción de las Naciones Unidas, así como información de cada uno de los ODS y de sus indicadores o métricas.

Goal		Show/Hide	List Operations	Expand Operations
GET	/v1/sdg/Goal/List		Returns the list of all Goals available	
GET	/v1/sdg/Goal/{goalCode}/Target/List		Returns the list of all Targets available for a Goal	
GET	/v1/sdg/Goal/{goalCode}/GeoAreas		Returns a list of geographies that have values on that Goal	
GET	/v1/sdg/Goal/{goalCode}/Dimensions		Returns a list of dimensions for a series	
GET	/v1/sdg/Goal/{goalCode}/Attributes		Returns a list of attributes for a series	
GET	/v1/sdg/Goal/Data		Returns a list of paginated observations	
POST	/v1/sdg/Goal/DataCSV		Returns a csv file	
POST	/v1/sdg/Goal/DataExcel		Returns an Excel file	
GET	/v1/sdg/Goal/PivotData		Returns a list of paginated observations pivoted by year	

Ilustración 5 Imagen que muestra parte de la API de las naciones Unidas.

Se ha dedicado mucho tiempo al estudio de esta API la cual cuenta con más de 40 servicios, debido a que los ODS son propuestos por la misma ONU, sería muy positivo contar con una gran cantidad de textos catalogados por ellos mismos, esto reduciría los fallos producidos por el incorrecto etiquetado de los textos.

Después de analizar los servicios propuestos, y probar muchos de ellos, hemos reducido la lista a 2 servicios de la API REST que fueron los más prometedores para recopilar grandes cantidades de información en formato texto:

GET	/v1/sdg/Series/List	Returns the list of all Series available
Implementation Notes Returns the list of all Series currently published in the latest release, you can use the allreleases flag to include archived data		

Ilustración 6 Imagen que muestra la descripción del método get v1/sdg/Series>List

La primera petición que vamos a mencionar es una petición GET, la cual devuelve un fichero con información sobre las ultimas publicaciones de la ONU, sin embargo, puede activarse un flag para que envíe todas las que hay archivadas.

A continuación, se muestra un extracto de la respuesta a la petición GET con el flag activado:

Response Body
[{ "goal": ["2"], "target": ["2.c"], "indicator": ["2.c.1"], "release": "2017.Q2.G.01", "code": "AG_FPA_MILLET", "description": "(Not In Use) Indicator of Food Price Anomalies (IFPA), millet", "uri": "/v1/sdg/Series/AG_FPA_MILLET" }, { "goal": ["15"], "target": ["15.c"], "indicator": ["15.c.1"], "release": "2017.Q2.G.01", "code": "AG_FPA_MILLET", "description": "(Not In Use) Indicator of Food Price Anomalies (IFPA), millet", "uri": "/v1/sdg/Series/AG_FPA_MILLET" }]

Ilustración 7 Imagen que muestra la respuesta a la petición get v1/sdg/Series>List

Como podemos observar se indica el ODS al que pertenece, en el atributo “goal”, se especifica incluso la meta y el indicador del que forma parte. Otros datos incluyen: fecha de lanzamiento, código, descripción y uri.

El código servirá como parámetro en las siguientes peticiones, para referencias un artículo concreto.

POST	/v1/sdg/Series/DataCSV	Returns a csv file
------	------------------------	--------------------

Ilustración 8 Imagen que muestra la petición POST /v1/sdg/Series/DataCSV

Como muestra la descripción del servicio, esta herramienta lo que hace es devolver un archivo en formato CSV, la petición necesita de un “código de serie”, este código es el que hemos obtenido de la petición GET anteriormente descrita,

así que nos disponemos a realizar una petición POST, con un código como parámetro.

Tras la petición realizada con el código “DC_ODA_BDVDL” descargamos el CSV generado y lo visualizamos:

Ilustración 9 Imagen que muestra el contenido del CSV devuelto por la petición POST

Tras analizar el texto descubrimos que es una tabla con datos sobre el gasto destinado por cada país a las distintas metas de un ODS concreto.

Queda claro que la información obtenida de esta forma no serviría para entrenar un modelo de tópicos, ya que la grandísima mayoría de las palabras que formarían los tokens se repiten en cada línea, cambiando solo el año y el país o la fuente o la métrica de una línea a otra.

Intentamos realizar más peticiones para verificar que lo observado en el último caso no es un evento aislado, realizamos más de 20 peticiones usando diferentes códigos y los resultados siempre fueron similares, por lo que decidimos abandonar esta línea de investigación.

4.2.2 Naciones Unidas Biblioteca Digital⁹



Ilustración 10 Imagen que muestra el buscador de la Biblioteca Digital de las Naciones Unidas

Biblioteca digital de las Naciones unidas, más de un millón de textos en total, aglutina documentos de diversas fuentes, la mayoría producidos por la misma Organización de las naciones unidas.

Tiene licencia Copyleft y permite la descarga directamente, sin embargo, no deja filtrar por ODS, ha de realizarse una búsqueda utilizando palabras claves, una opción sugerida por el personal de atención de este portal para solventar este inconveniente, reside en utilizar la taxonomía publicada por ellos mismos a la hora de realizar la búsqueda.

Sustainable Development Goals Taxonomy

- Goal 1: No poverty End poverty in all its forms everywhere
 - Target 1.1: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day
 - Indicator 1.1.1: Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)
 - Target 1.2: By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
 - Target 1.3: Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable
 - Target 1.4: By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance
 - Target 1.5: By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters
 - Target 1.a: Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and policies to end poverty in all its dimensions
 - Target 1.b: Create sound policy frameworks at the national, regional and international levels, based on pro-poor and gender-sensitive development strategies, to support accelerated investment in poverty eradication actions

Ilustración 11 Imagen que muestra la Taxonomía propuesta por las Naciones Unidas

Sin embargo, este método acarrea dos problemas, el primero que, dentro de los textos filtrados, habría una cantidad de textos no determinada, que, a pesar de incluir términos de la taxonomía relativa al objetivo buscado, trata de otro de los objetivos, o ninguno de ellos.

El otro problema es que, al utilizar términos de búsqueda específicos, inevitablemente el corpus mostrará una sobrerepresentación de dichos términos de búsqueda.

Debido a estos inconvenientes de este portal, hemos decidido continuar buscando otros portales de datos.

4.2.3 UN-iLibrary

Página de las naciones unidas que recopila artículos de diversas fuentes, permite filtrar los artículos por palabras clave, pero también según el ODS al que pertenecen, esto sumado a que cuenta con una gran cantidad de textos, lo hace una fuente de documentos idónea para este proyecto, si no fuese por el hecho de que los textos no son de libre acceso, sino que más bien se tratan de revistas y otros documentos de pago.

Decidimos contactar con el servicio de atención de este portal para verificar si hubiese alguna forma de utilizar sus datos con fines educativos, ante la negativa a esta petición, decidimos descartar la fuente.

The screenshot shows the advanced search page of the UN-iLibrary website. At the top, there is a header with the UN-iLibrary logo, a search bar, and language selection (ES). Below the header, the URL 'Inicio / Busqueda avanzada' is visible. The main search area is titled 'Busqueda avanzada'. It contains two sets of input fields for keywords/phrases, with dropdown menus for 'in' and 'Todos'. A '+' button is available to add more search terms. There is also an option to exclude full text. Below these are sections for 'BUSCAR ENTRE ESTAS FECHAS:' (Search between these dates) and 'ORDENAR RESULTADOS POR:' (Order results by relevance). The 'TIPO DE CONTENIDO:' dropdown is set to 'Todos'. A 'FILTER BY SDG:' section lists 'SDG 1: No Poverty', 'SDG 2: Zero Hunger', 'SDG 3: Good Health and Well-Being', and 'SDG 4: Quality Education'. To the right of the search form, there is a sidebar titled 'Access Key' showing 'UN Titles Subscribed' and a 'COVID-19 Reports' section featuring a circular graphic and a 'Learn More' button. A note at the bottom right states: 'The United Nations provides new information on COVID-19 as it becomes available, and we are sharing the latest reports and policy briefs here on UN iLibrary.'

Ilustración 12 Imagen que muestra el buscador avanzado de UN-iLibrary

4.3 Corresponsables¹⁰

Página de noticias de carácter internacional, especialmente en países hispanoparlantes, **permite filtrar por ODS** y cuenta con más de 10.000 artículos, documentos de libre acceso, sin necesidad de suscripción ni registro.

The screenshot shows the Corresponsables website's search interface. At the top, there are links for W3C, WCAG 2.0, Acceso / Registro, and several language options: ARGENTINA, CHILE, COLOMBIA, ECUADOR, ESPAÑA, MÉXICO, PERÚ, and INGLÉS. Below this is a navigation bar with links for Noticias, Opinión, Publicaciones, Agenda, TV, and Recursos. The page title is "Actualidad". A breadcrumb trail indicates the user is at "Inicio / Actualidad". The main search form includes fields for "Que contenga" (What contains), "Grupos de interés" (Interest groups), "ISO 26000", "ODS" (with a dropdown menu showing "ODS 11. Ciudades y Comunidades Sostenibles"), "Sector", "Desde" (From), "Hasta" (To), and an "Aplicar" (Apply) button. Two news items are displayed below the search form:

- ODS 11. Ciudades y Comunidades Sostenibles** (highlighted with a red box).
Thumbnail: A photo of a modern interior room.
Text: "ODS11. Los fondos europeos darán tres grandes oportunidades a la rehabilitación de viviendas".
Details: Published on 30/04/2021, 11:11, 11:11 Noticia.
Buttons: "Ver más".
- PREMIOS NACIONALES DE MOVILIDAD** (with a small logo).
Thumbnail: A green graphic with the text "PREMIOS NACIONALES DE MOVILIDAD".
Text: "ODS11. Gala de entrega de los Premios Nacionales de Movilidad".
Details: Published on 30/04/2021, 11:11, 11:11 Noticia.
Buttons: "Ver más".

Ilustración 13 Imagen que muestra la página de Corresponsables, en su sección de actualidad

Como inconveniente, los textos no tienen opción de ser descargados de manera directa, por lo que han de ser extraídos desde la misma página web, a través de web scraping.

Los portales de noticias generaron gran interés en este proyecto, debido a gran cantidad de documentos que generan, sin embargo, no encontramos ninguno con las características de Corresponsables, muchos de ellos requieren de pago para acceder a todos sus archivos, y no se encontró ningún otro que permitiese filtrar por ODS.

4.4 SDG PATHFINDER¹¹



Ilustración 14 Imagen que muestra la interfaz web de la página SDG PATHFINDER

Esta web **deja filtrar por ODS**, y cuenta con una buena cantidad de textos, sin embargo, los textos son de difícil recopilación, al estar los documentos contenidos en distintos formatos (PDF, Diversas páginas web, cada una con una estructura HTML distinta.)

Desarrollado por la OCDE “Organización para la Cooperación y el Desarrollo Económicos” una importante organización que colabora estrechamente con la ONU lo que le convierte en una importante fuente de información a considerar.

Textos redactados en inglés y cantidad de textos muy dispar entre distintas ODS, habiendo algunos ODS con pocos documentos, el filtrado de documentos se realiza

Por último, algunos de los documentos en este portal, están marcados como pertenecientes a más de un ODS, y aunque esto sea lógico, dada la cercanía de algunos de los objetivos de desarrollo, creemos que es mejor que los textos de entrenamiento solo pertenezcan a un ODS.

2 ZERO HUNGER



Zero hunger

A profound change of the global food and agriculture system is needed. Today we produce more than enough food to nourish the world's population. Yet in 2018, over 815 million people (almost 11% of the world's population) – mostly from developing

Show more ▾



Ilustración 15 Imagen que muestra el mapa de tópicos, con los tópicos para el Objetivo de "Hambre Cero" resaltados.

Cuenta también con un mapa de tópicos, donde puedes ver la cercanía de cada tópico con cada uno de los ODS y el volumen relativo de documentos que contienen el tópico.



Ilustración 16 Imagen que muestra el Mapa de Tokens completo, de la página SDG PATHFINDER.

5 Herramientas de Desarrollo

5.1 Python

Python es un lenguaje de programación orientado a objetos, es el lenguaje de programación que más creció en uso en el último año, principalmente por su facilidad de uso y la gran cantidad de librerías que tiene sobre ML y WebScraping, Es precisamente por estas librerías por lo que encaja con el objetivo de este proyecto, además es un lenguaje opensource, nos hemos decantado por Python como lenguaje principal en este proyecto.

5.2 librAlry

Herramienta desarrollada por un equipo multidisciplinar de la Universidad Politécnica de Madrid, Carlos Badenes-Olmedo, Jose Luis Redondo y Oscar Corcho. En la web la describen como: "Conjunto de servicios software integrados, centrados en facilitar la exploración de grandes colecciones de documentos"¹².

Combinas técnicas de NLP, ML y recursos semánticos para exponer APIs a través de interfaces Restful, aun que tiene diversos casos de uso contemplados, en este proyecto vamos a centrarnos en la funcionalidad que analiza colecciones de documentos para descubrir temas ocultos en sus textos y crear modelos de aprendizaje¹³ basado en tópicos.

5.3 Docker

Docker es un contenedor que ofrece un modelo de implementación basado en la creación de imágenes, esto permite separar los procesos y compartirlos en varios entornos, de modo que puedan ejecutarse de manera independiente¹⁴. Docker está implementado sobre el kernel de Linux y en este proyecto lo usaremos para ejecutar de esta forma el servicio de librAlry el cual crea imágenes en Docker.

5.4 GitHub

Controlador de versiones más popular, se ha utilizado para mantener un registro de los cambios generados tanto del código generado como de la memoria.

Enlace a para acceder al repositorio con todo el código desarrollado durante el proyecto y las últimas versiones de la memoria y del corpus es: https://github.com/Julen-Perez/TFG_NLP_ODS.

5.5 Minería de textos

La minería de textos define la capacidad de encontrar patrones ocultos y conocimientos implícitos en grandes corpus de documentos. Por lo general en la minería de datos, la extracción no se realiza de corpus ordenados, donde la información es fácilmente extraíble, cosa que, si pasa en la minería de datos¹⁵, esto añade un elemento de dificultad a la extracción de dicha información lo que implica añadir el paso de tener que estructurar la información, previo análisis.

Una parte importante de la minería de datos es el área especializada en la exploración de material encontrado en Internet, que como ya hemos argumentado en este proyecto, se trata de una de las mayores fuentes de información de la actualidad, a esta modalidad se le denomina Web Mining.

La Web mining requiere tres pasos para el correcto procesado de la información.

- Definir los textos relevantes. La información que vamos a obtener va a ser solamente tan valiosa como lo sean los textos que procesar, es por eso por lo que esta fase es vital para todo proceso de minería de textos. Hay que definir el conjunto de textos que son relevantes y la fuente de la que se van a obtener. Esto puede tener un componente subjetivo si se hace de forma manual, dependiendo de qué características del texto se denotan como más relevantes, por ejemplo, si el investigador prioriza tener una gran cantidad de textos, frente a una fuente que tenga menos textos, pero de mayor longitud. El investigador también puede apoyarse en índices y motores de búsqueda para realizar una elección.
- Recopilación de los documentos. Etapa en la que se trae la información de la fuente web a un formato texto, para que pueda posteriormente ser procesada. Este paso intermedio suele almacenar la información, por ejemplo, en archivos para su posterior procesado. Como hemos comentado antes, las técnicas de Minería de datos o Data mining necesitan que la información esté estructurada, es por esto por lo que este paso es de vital importancia.
- Data Mining; Esta es la etapa en la que se descubren patrones y características ocultas en el corpus original. Se establecen patrones y se sacan conclusiones.

5.6 Web Scraping

Dentro de la etapa de extracción de documentos en Web Mining, se incluyen diversos métodos de obtener dichos documentos, la técnica utilizada va a variar a depender tanto de la preferencia y elección del investigador, como del formato inicial en el que se encuentre la información, los dos más destacables por su uso en este proyecto son:

- Peticiones API REST: API conocido por sus siglas en inglés Application Programming interface, es una herramienta para compartir contenido e información¹⁶, en el ámbito de data mining, se usan para extraer información de una base de datos, de una manera más estable y programática que la que permite una interfaz de usuario regular¹⁷. El inconveniente de este método es que solo es posible cuando la página de la que se desea extraer información cuenta con una API, y esta debe ser abierta para

permitir el acceso libre. A su vez, la información contenida por la API ha de ser de utilidad para el investigador.

- Web scraping: Una alternativa para cuando la web no cuenta con una API libre consiste en la técnica conocida como Web scraping, esta técnica consiste en extraer directamente de la página la información que esta contiene. Esta técnica sin embargo cuenta con dificultades, ya que requiere un estudio de la página web y de su estructura por parte del investigador, en especial si se pretende automatizar el proceso de extracción.

5.7 Beautiful Soup

Librería de Python que permite explorar y extraer información de los archivos HTML y XML de una página web, posibilita adquirir el código fuente de la página y filtrar lo que el investigador considere de relevancia.¹⁸

A la hora de realizar este proyecto se ha utilizado la versión 4.9.3.

5.8 Web Drivers

Un web driver es una herramienta utilizada para automatizar procesos web, a través de navegadores. En español controlador, permite navegar a través de la web, imitando el comportamiento de un usuario¹⁹, también permite interactuar de manera automática con el navegador, enviado caracteres para que escriba en los campos adaptados para ello, o incluso permite realizar ejecuciones de JavaScript.

Dado que los web drivers son específicos de cada navegador, en este proyecto nuestro controlador de elección será Chromium, el driver específico para Google Chrome. Para hacer uso de dicho controlador, usaremos la librería Selenium²⁰ de Python, muy popular entre desarrolladores para el automatismo de procesos web.

5.9 Procesamiento de lenguaje natural

El Procesamiento del Lenguaje Natural(PLN), es una herramienta multidisciplinaria que combina la Inteligencia Artificial y Lingüística, se centra en la comunicación entre humanos y computadores, a través del procesamiento del lenguaje natural que utilizamos las personas, haciendo que la máquina interprete el lenguaje, y pueda incluso producir lenguaje de forma autónoma²¹.

El PLN aspira a modelar los mecanismos cognitivos ocultos en la creación y el entendimiento de las lenguas humanas. Dichas lenguas pueden expresarse de diversas maneras, pasando del lenguaje hablado, al lenguaje de signos o el lenguaje escrito, en este proyecto nos centraremos en el lenguaje en formato texto, en el que se encuentra la mayor parte de los datos almacenados.

Con el paso de los años y la introducción de nuevas tecnologías, el PLN ha pasado por tres fases, cada una de ellas caracterizada por una manera de enfocar el problema²²:

- Sistemas basados en reglas, marco los inicios del PLN, se basa en la creencia de que el comportamiento del lenguaje es estático, definiendo unas reglas concretas que son capaces de definir el espectro del lenguaje natural. Con el avance de los años, este enfoque ha quedado prácticamente en desuso.
- Machine Learning, enfoque más genérico, se centra en analizar características del lenguaje natural, a través del procesador de textos, para obtener modelos estadísticos, como la ocurrencia de las palabras.
- Deep Learning última tendencia y la que más se acerca a imitar el lenguaje humano, nace gracias a los avances en el procesamiento de cálculo y a tener grandes cantidades de datos a analizar, no necesita de intervención humana más allá de alimentarle los datos, el modelo aprende a través de una red neuronal características del lenguaje.

Los usos del PLN son diversos y han dado lugar distintas líneas de investigación, cada una enfocada a múltiples problemas a resolver. Algunas de estas líneas de investigación son²³:

- Análisis de la información y síntesis de un resumen.
- Búsqueda de respuesta a una pregunta.
- Identificación de conceptos claves y su relación con entidades.
- Técnicas automáticas de clasificación de textos.
- Traducción.
- Reconocimiento de caracteres.
- Reconocimiento automático del discurso hablado.
- Etiquetado morfológico, sintáctico y semántico.
- Análisis de sentimientos.

Este trabajo pertenece al área de investigación de Técnicas automáticas de clasificación de textos y como ya hemos comentado, haremos uso del modelo Labelled LDA como modelo de tópicos.

[5.10 Labelled LDA²⁴](#)

Los modelos de tópicos son modelos estadísticos que se utilizan para descubrir temas que sean representativos de una colección de textos, esta metodología ha ido creciendo en popularidad en los últimos años.

El modelo de tópicos más extendido en la comunidad científica es el LDA, propuesto en el año 2002, se trata de un algoritmo no supervisado, que modela cada documento como una mezcla de tópicos. El modelo genera resúmenes en forma de términos, basándose en la recurrencia de cada palabra sobre cada tópico.

A pesar del éxito que tiene este modelo, tanto en la comunidad científica como a la hora de modelar documentos, carece de posibilidad de incorporar

documentos que ya estén categorizados en alguno de los tópicos, dificultando así su entrenamiento supervisado, aunque se contase con textos y tiempo suficiente para clasificar manualmente los textos, previamente a incorporarlos al modelo.

Para solventar esta carencia, nace el modelo Labelled LDA, este modelo probabilístico es una extensión del LDA, y al igual que este, modela cada documento como una suma de tópicos y genera un término para cada tópico. Sin embargo, y a diferencia del LDA, esta variante ofrece la posibilidad de incorporar elementos ya clasificados, para entrenar el modelado de los documentos.

6 Desarrollo

6.1 Fase 1 Selección de una fuente

La idea inicial consistía en utilizar Google académico para hacer búsquedas sobre cada uno de los ODS, y a pesar de las dificultades de este método, fue así como encontramos la página corresponsables.com, que a cada artículo le asigna uno de los ODS y le añade dicho ODS al nombre del artículo, por lo que era un resultado muy frecuente en las búsquedas de Google Académico.

Además de Google scholar, se intentó hacer búsquedas especializadas en portales de datos, los cuales son mayoritariamente de lengua inglesa, sin embargo, esto acarreaba un problema similar al descrito en la sección 4.4.2, donde las búsquedas estaban determinadas por los términos que se utilizan para hacer la búsqueda, corriendo así el riesgo de influenciar a la hora de crear los tokens.

Como vimos anteriormente, la API de la ONU sobre los ODS no nos aportó información relevante para el entrenamiento del modelo, al tratarse mayoritariamente de información repetitiva. Tampoco nos sirvió el portal de documentos ni de artículos de la misma ONU.

Después de revisar múltiples fuentes, se decidió usar Corresponsables.com, al ser una fuente de información relevante para los temas, de fácil acceso y ya clasificada en cada uno de los ODS. Los temas son de actualidad al tratarse de un portal de noticias, a pesar de no poder ser extraídas de manera directa, como por petición a una API o a través de descargas, cuenta con una gran cantidad de textos en castellano, por lo que ha sido la elegida como fuente de información para los textos catalogados en este proyecto.

A pesar de todos estos aspectos positivos, el portal no cuenta expresamente con una licencia de datos abiertos, de hecho, no especifica el tipo de licencia, debido a esto los textos entrarian dentro de la licencia de Copyright, que es la predeterminada. Tras intentar contactar con la organización en múltiples ocasiones y por todos los medios disponibles en su página web, decidimos seguir adelante y utilizar los textos recopilados en su portal, hicimos esto amparándonos en las excepciones al copyright que comentamos en la sección de “Licencias”.

En ningún momento pretendemos vulnerar los derechos de autor de ninguna página, el motivo del uso de las fuentes es puramente académico y nunca con ánimo de lucro. Las obras utilizadas no son mostradas, solo son utilizadas en el entrenamiento del modelo de tópicos. No se pretende hacer competencia directa o indirecta a la organización Corresponsables.

En resumen, la fuente escogida para construir el corpus de texto fue el de corresponsables.

6.2 Fase 2 Recuperación de datos.

Para solventar el problema del acceso a los datos, se han utilizado las tecnologías de web scraping, descritas en el apartado “Herramientas utilizadas”. Se desarrolló un programa Python²⁵ que navegue a través de la página web, y automatizase el proceso de recogida de los documentos:

Lo primero que tenemos que hacer es realizar los import de las librerías que vamos a utilizar:

```
from selenium import webdriver
from time import sleep
from bs4 import BeautifulSoup as soup # HTML data structure
from webdriver_manager.chrome import ChromeDriverManager
import urllib.request
```

Ilustración 17 Imagen que muestra los imports realizados para la obtención de textos.

Se pueden apreciar las librerías webdriver y BeautifulSoup de las que hablamos en el apartado de “Herramientas utilizadas”. También se puede apreciar la librería time, de la cual utilizaremos la función sleep, para introducir un retardo entre acciones del webdriver, para que dé tiempo a completar la acción en curso, antes de dar paso a una acción posterior.

Una vez importados las librerías, debemos inicializar el driver, y guardarlo en el atributo que llamaremos driver, esto lo hacemos en el constructor de Python `__init__` y en este caso no recibe parámetros.

```
class ODS_BOT_OBTENER_ENLACES():
    def __init__(self):
        self.driver = webdriver.Chrome(ChromeDriverManager().install())
    def parser(self,page_url):
        uClient = urllib.request.urlopen(page_url)
        page_soup = soup(uClient.read(), "html.parser")
        uClient.close()
        containers = page_soup.findAll("div", {"class": "views-field views-field-field-body"})
        cleantext = soup(str(containers[0]),'html.parser').text.replace("\n", " ")
        return cleantext
    def obtener_enlaces(self,numero_ods):
        lista_ods = self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/form/div/div/div/div[4]/div/div/select')
        lista_ods.click()
        sleep(1)
        escoger_filtro_ods = self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/form/div/div/div/div[4]/div/div/select/option[{}].format(numero_ods+1)')
        escoger_filtro_ods.click()
        sleep(2)
        aplicar = self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div[1]/form/div/div/div[8]/input')
        aplicar.click()
        sleep(2)
        i=1
        articulo=[]
        while True:
            try:
                mostrar_mas= self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/u1/li/a')
                mostrar_mas.click()
                print(i)
                sleep(4)
                i+=15
            except Exception:
                break
        for x in range(i):
            try:
                articulo.append(self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div[2]/div[{}]/span[{}].format(x+1)).get_attribute("href"))
            except Exception:
                continue
        print(len(articulo)-1)
        print(len(articulo))
        return articulo
bot = ODS_BOT_OBTENER_ENLACES()
bot.driver.get('https://www.corresponsables.com/actualidad')
sleep(4)
filename = "texts_prueba.csv"
j=1001;
with open(filename, "w", encoding="utf-8") as f:
    for c in range(17):
        enlaces=bot.obtener_enlaces(c+1)
        textos=[]
        for enlace in enlaces:
            try:
                sol=bot.parser(enlace)
                f.write(str(j)+" ; "+ "ODS"+str(c+1)+"; "+sol+"\n")
                j=j+1
            except Exception:
                continue
        print("exito ODS"+str(c+1))
f.close()
```

Ilustración 18 Imagen que muestra el código desarrollado en Python para la obtención de textos de la página Corresponsables.

Creamos una instancia de la clase ODS_BOT_OBTENER_ENLACES a la que llamaremos bot. Y haremos que el driver actúe en la dirección de la página de actualidad de corresponsables, donde están contenidos todos los documentos.

Tras abrir en modo lectura el fichero donde vamos a volcar el texto extraído, pasaremos a llamar 17 veces, una vez por cada ODS, a la función obtener_enlaces la cual interactuará con la interfaz web y filtrará cada uno de los objetivos de desarrollo, para así poder recopilar todos los enlaces web a todos los artículos existentes de cada tema filtrado.

El primer problema que nos encontramos fue filtrar de manera automática el contenido en la sección de actualidad de la página web, dado que se trata de una dirección url común para todas las noticias, se requería la interacción con el desplegable y con el botón para filtrar cada uno de los ODS. Haciendo uso de la funcionalidad find_element_by_xpath(p) del driver, siendo p el xpath que se ha obtenido del inspeccionado del elemento a encontrar en la página web. Es así como somos capaces de encontrar y desplegar la lista desplegable de ODS(1), luego elegir el ODS por el que hayamos hecho la llamada de obtener_enlaces(2), y por último aplicar el filtro seleccionado(3), para que solo muestre los artículos del ODS que nos interesa.

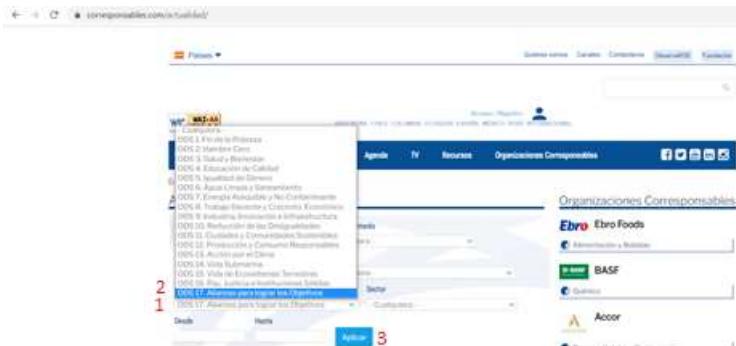


Ilustración 19 Imagen que muestra la interfaz web de Corresponsables, con el menú de Objetivos de desarrollo desplegado.

Otro elemento que necesitó de interacción fue, una vez elegido el ODS correspondiente, que la página liste todos los artículos disponibles, ya que, por defecto, solo muestra los últimos 15 artículos añadidos, y al pulsar el botón de mostrar más, solo muestra otros 15. Otro problema es que, a priori, no sabemos cuántos artículos contiene la página de cada tópico, y que no todos los tópicos tienen el mismo número de documentos. Es por esto por lo que forzamos al driver a pulsar el botón de mostrar más, hasta que no sea posible pulsarlo de nuevo, momento en el cual da una excepción y sabremos que hemos llegado hasta el último artículo disponible.

Puede verse también, que tras cada acción de “mostrar_mas”, hay una pausa de 4 segundos, que inicialmente era menor, pero se ha ido ampliando para evitar errores.

Tras tener listados todos los artículos de un ODS concreto, se añaden a un vector para ser devueltos a la función main, y ser allí tratados y enviados al CSV, el cual se usará para el entrenamiento del modelo de tópicos. Una cuestión que ha facilitado la recopilación de los enlaces a los artículos es que todos comparten la misma dirección xpath, exceptuando una de las posiciones de uno

de sus div[] que incrementa de uno a uno entre artículos, siendo el número 1, el último artículo añadido a la página.

Una cosa para tener en cuenta es que, en ocasiones, nos encontramos con una noticia que no tenía href a un artículo, esto puede ser por que se tratase de un anuncio o un video. En estos casos el driver no encontraría el xpath y daría un error, es por eso que se ha añadido una excepción que simplemente continua a la siguiente iteración del bucle.

Una vez obtenidos todos los enlaces de los artículos del ODS, devolvemos el vector con los enlaces, y por cada enlace, hacemos una llamada a la función parser(self,page_url), el objetivo de esta función es recoger el texto contenido en cada enlace que se le pasa como parámetro.

```
def parser(self,page_url):
    uClient = urllib.request.urlopen(page_url)
    page_soup = soup(uClient.read(), "html.parser")
    uClient.close()
    containers = page_soup.findAll("div", {"class": "views-field views-field-field-body"})
    cleantext = soup(str(containers[0]),'html.parser').text.replace("\n", " ")
    return cleantext
```

Ilustración 20 Imagen que muestra el método parser, empleado para obtener el texto de la URL indicada.

Lo primero que necesitamos hacer es abrir la url que hemos obtenido, para esto hacemos un ulropen request de la librería urllib, pasándole de parámetro la url en cuestión. Posteriormente obtener un objeto de tipo soup, teniendo como primer parámetro la página abierta por la petición previamente descrita, y como segundo parámetro “html.parser”, para que devuelva la estructura html de la página. Tras esto podemos cerrar ya la página haciendo uso de la función close().

Después de esto necesitamos saber dónde, de todo el esqueleto html está el texto que necesitamos extraer, obviando el resto del contenido de la página que no nos resulta de utilidad. Tras analizar la estructura, descubrimos que la totalidad del texto a extraer se encuentra en la misma clase, class="views-field views-field-field-body". Una vez sabemos esto, solo hay que extraer el texto de esa clase con la librería soup.

Además de recoger el texto, es necesario que esté en el formato correcto, para ello se han eliminado los saltos de línea y se han introducido en el CSV que será subido a librAIry, el CSV consiste en 3 columnas de datos, la primera un identificador único del documento, la segunda una etiqueta, mostrando a que tópico pertenece, en este modelo será uno por ODS, y, por último, una columna con el texto del documento, todo ello en la misma línea.

Además de eliminar los saltos de línea, hemos realizado otras modificaciones leves al texto extraído, como entrecollar el texto o eliminar algunos signos de puntuación, aunque la herramienta librAIry detecta y realiza muchos de estos cambios de forma automática.

Tampoco ha sido necesario eliminar palabras de “relleno” que no aportan información del texto como podrían ser “de” o “y”, ya que librAIry las detecta y las ignora a la hora de formalizar el modelo.

6.3 Fase 1 y 2 bis

El fichero mostrado en la ilustración 18 sirve para extraer todos los artículos de la página de Corresponsables hasta la fecha, consecuentemente lo lógico sería

hacer una única extracción de los textos, en una fecha lo más tardía posible, y así se planteó en un primer momento.

Sin embargo, en el transcurso de este proyecto nos percatamos de que, la página además de introducir nuevos artículos también eliminaba artículos de la web, la cadencia de dichas cribas no ha podido ser determinada, tampoco hemos podido identificar el criterio de que artículos pasaban a ser eliminados, dado que no era una cuestión temporal, al permanecer artículos de 2017 y eliminar algunos posteriores.

Ante esta situación de incertidumbre, decidimos hacer una primera copia del contenido de la página web, con todos los artículos que hubiese en ese momento, dicha copia se realizó el día 02/04/2021.

Para complementar el corpus realizado tras esa fecha, creamos aparte un programa Python que trajese todos los artículos cuya fecha de subida sea posterior a la indicada, pudiendo así extraer todos los artículos que se añadan nuevos en la página y no estén aún en el corpus. La fecha de la última modificación ha de ser indicada en el campo “fecha_ultima_actualizacion”, de modo que el programa solo recoja artículos posteriores esta fecha.

```
def actualizar_enlaces(self,a,fecha_ultima_actualizacion):
    lista_ods= self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/form/div/div/div/div[4]/div/div/select')
    lista_ods.click()
    sleep(1)
    escoger_filtro_ods = self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/form/div/div/div/div[4]/div/div/select/option[{}].format(a+1)')
    escoger_filtro_ods.click()
    sleep(2)
    aplicar= self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/form/div/div/div/div[8]/input')
    aplicar.click()
    sleep(2)
    i=0
    articulo=[]
    while True:
        if i<0 and i>15==0:
            mostran_mas= self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[1]/ul/li/a')
            mostran_mas.click()
            sleep(4)
        stri= bot.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[2]/div[{}]/div[3]/span'.format(i+1)).text
        fecha_noticia = datetime.strptime(stri, '%d/%m/%Y').date()
        if fecha_ultima_actualizacion > fecha_noticia:
            try:
                articulo.append(self.driver.find_element_by_xpath('/html/body/div[3]/div[4]/div[3]/div[1]/section/div[2]/div/div/div/div/div/div[2]/div[{}]/span/a'.format(i+1)).get_attribute("href"))
                i+=1
            except Exception:
                i+=1
                continue
        else:
            break
    return articulo

bot = ODS_BOT_ACTUALIZAR_ENLACES()
bot.driver.get('https://www.corresponsables.com/actualidad')
sleep(4)
filename="texts_prueba.csv"
j=11058;
p=11058;
fecha_ultima_actualizacion=datetime.strptime('16/04/2021', '%d/%m/%Y').date()
numero_enlaces=[]
with open(filename, "a+", encoding="utf-8") as f:
    for c in range(17):
        enlaces=bot.actualizar_enlaces(c+1,fecha_ultima_actualizacion)
        numero_enlaces.append(len(enlaces))
    for enlace in enlaces:
        try:
            sol=bot.parser(enlace).replace("\n", " ")
            f.write(str(j)+";"+ODS[{}].format(c+1)+";;"+sol+"\n")
            j+=1
        except Exception:
            numero_enlaces[c]-=1
            continue
        print("Añadidos {}".format(numero_enlaces[c])+"enlaces para la ODS {}".format(c+1))
    f.close()
    print("Añadidos un total de: {}".format(j-p))
exit()
```

Ilustración 21 Imagen que muestra el código desarrollado para la actualización incremental del corpus, con textos de Corresponsables.

El código desarrollado para actualizar los nuevos enlaces es muy similar al usado para la extracción inicial, exceptuando, que ahora itera hasta que la fecha de un artículo sea menor a la indicada (fecha de última actualización). Otro elemento diferente es que hay que indicarle también el id del siguiente elemento a añadir (incrementando uno al último id existente).

De esta forma logramos incorporar más de 600 textos nuevos, en sucesivas actualizaciones, los textos se añaden directamente al CSV, cuyo nombre ha de

estar indicado en la variable “filename”, esto es posible gracias a la función append, para lo cual hemos tenido que abrir el fichero en modo “a+” como muestra la imagen superior.

Las imágenes a continuación muestran la respuesta a dos ejecuciones independientes del programa, en ellas se muestra el número de textos añadidos por cada uno de los ODS, y el número total de textos añadidos. Este programa en Python permite la actualización incremental del corpus de manera indefinida, siempre y cuando la página de corresponsables siga añadiendo artículos.

Añadidos 4 enlaces para la ODS 1	Añadidos 4 enlaces para la ODS 1
Añadidos 2 enlaces para la ODS 2	Añadidos 7 enlaces para la ODS 2
Añadidos 13 enlaces para la ODS 3	Añadidos 42 enlaces para la ODS 3
Añadidos 8 enlaces para la ODS 4	Añadidos 13 enlaces para la ODS 4
Añadidos 5 enlaces para la ODS 5	Añadidos 10 enlaces para la ODS 5
Añadidos 0 enlaces para la ODS 6	Añadidos 1 enlaces para la ODS 6
Añadidos 6 enlaces para la ODS 7	Añadidos 14 enlaces para la ODS 7
Añadidos 13 enlaces para la ODS 8	Añadidos 41 enlaces para la ODS 8
Añadidos 5 enlaces para la ODS 9	Añadidos 18 enlaces para la ODS 9
Añadidos 3 enlaces para la ODS 10	Añadidos 25 enlaces para la ODS 10
Añadidos 7 enlaces para la ODS 11	Añadidos 7 enlaces para la ODS 11
Añadidos 8 enlaces para la ODS 12	Añadidos 41 enlaces para la ODS 12
Añadidos 8 enlaces para la ODS 13	Añadidos 33 enlaces para la ODS 13
Añadidos 0 enlaces para la ODS 14	Añadidos 2 enlaces para la ODS 14
Añadidos 2 enlaces para la ODS 15	Añadidos 9 enlaces para la ODS 15
Añadidos 5 enlaces para la ODS 16	Añadidos 22 enlaces para la ODS 16
Añadidos 12 enlaces para la ODS 17	Añadidos 19 enlaces para la ODS 17
Añadidos un total de: 101	Añadidos un total de: 308

Ilustración 22 Imágenes que muestran distintas actualizaciones del corpus al ejecutar el código de actualizar_enlaces.

A continuación, se muestra el número total de textos recuperados de cada uno de los ODS usando tanto la primera recuperación de textos como las sucesivas actualizaciones:

ODS	Número de textos
1	328
2	321
3	1503
4	507
5	429
6	112
7	341
8	1257
9	413
10	801
11	434
12	679
13	775
14	77
15	47
16	782
17	1249
Total	10055

Ilustración 23 Tabla con el número de textos total recopilados por cada Objetivo de Desarrollo Sostenible.

6.4 Fase 3: Extracción

Una vez contamos con el corpus de entrenamiento, nos disponemos a generar el modelo de tópicos. Como hemos comentado esto lo haremos usando la herramienta librAIry y para ello necesitamos arrancar el Docker, y posteriormente realizar una petición a la API, a la dirección /topics.

Esta petición POST requiere de un JSON que va a aportar datos al clasificador, los que vamos a usar en este proyecto son:

- name: el nombre del modelo
- description: una descripción del modelo
- contactEmail: un email de contacto, para informar al usuario cuando haya terminado el entrenamiento
- version: La versión del modelo
- docker: Información sobre el Docker en el que se va a depositar la imagen una vez termine el entrenamiento.
- dataSource: Información relativa al fichero que contiene los documentos, como las cabeceras y el nombre del fichero.

- format: El formato en el que se encuentra el fichero
- offset: posición desde la cual se quiere empezar a leer el fichero.
- size: número de líneas que se quieren leer del fichero (-1 indica que todas)
- url: dirección en la que se encuentra el fichero

Además de todos estos parámetros, se pueden incluir otros como el idioma o el número de tópicos, sin embargo, librAIRy tiene detección automática de lenguaje, y al tratarse de un modelo Labelled LDA, que ya cuenta con las etiquetas, infiere el número de tópicos del número de etiquetas distintas que se le proporcionan.

A pesar de tener el fichero en local, librAIRy requiere que se le proporcione una url desde la cual pueda acceder al fichero. Inicialmente intentamos usar plataformas de almacenamiento en la nube como Dropbox o Google Driver, sin embargo, tratarse de un corpus con muchos documentos, Dropbox cerraba la conexión antes de que pudiese completarse el entrenamiento, es por eso que finalmente optamos por realizar una subida del documento a Solr, desde donde poder acceder al CSV. Para subir el fichero CSV a Solr realizamos otro POST, en la API de librAIRy, en la extensión /documents.

```
{
  "contactEmail": "julen.perez.alvarez@gmail.com",
  "dataSink": {
    "format": "SOLR_CORE",
    "url": "http://librairy-repo:8983/solr/documents"
  },
  "dataSource": {
    "name": "texts_prueba_ODS",
    "dataFields": {
      "id": "0",
      "name": "0",
      "labels": ["1"],
      "text": ["2"]
    },
    "filter": "",
    "format": "CSV",
    "offset": 1,
    "size": -1,
    "url": "/librairy/batch/texts_prueba_ODS.csv"
  }
}
```

Ilustración 24 Imagen que muestra la estructura del JSON empleado para la petición de añadir el corpus a SOLR a través de librAIRy.

Una vez indexado el documento, ya podemos realizar la petición a /topics, indicando en el campo url la dirección de solr donde está el documento.

```
{
  "name": "ODS_Classifier_Model",
  "description": "topic model to classify ODS texts",
  "contactEmail": "julen.perez.alvarez@gmail.com",
  "version": "1.0",
  "docker": {
    "email": "julen.perez.alvarez@gmail.com",
    "password": "[%;rWqR_cy.Q.j7",
    "repository": "julenperez/nlp_ods1",
    "user": "julenperez"
  },
  "dataSource": {
    "name": "texts_prueba_ODS",
    "dataFields": {
      "id": "id",
      "labels": [
        "labels_t"
      ],
      "text": [
        "txt_t"
      ]
    },
    "format": "SOLR_CORE",
    "offset": 0,
    "size": -1,
    "url": "http://librairy-repo:8983/solr/documents"
  }
}
```

Tras generar la imagen en Docker, podemos acceder a ella con el siguiente comando:

```
docker run -it --rm -p 8585:7777 julenperez/nlp_ods1:1.0
```

De esta forma Podemos acceder a la API creada por librAIry al construir el clasificador, estará en el puerto 8585 siendo la dirección “<http://localhost:8585/api.html>”. Dicha API contamos con diferentes servicios:

Ilustración 25 Imagen que muestra el JSON utilizado para la generación del modelo de tópicos.

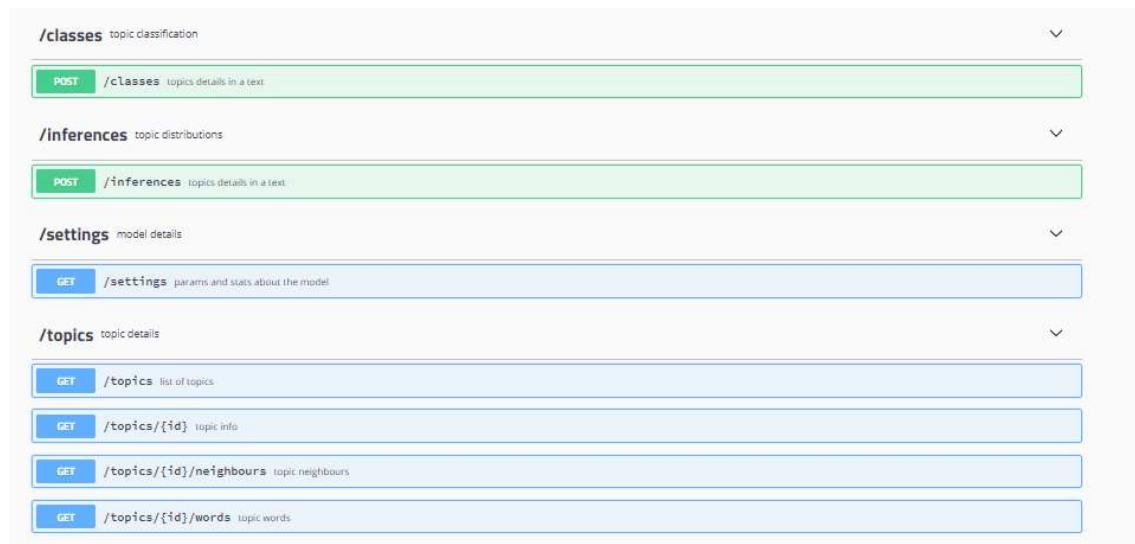


Ilustración 26 Imagen que muestra la API generada por librAIry.

De estos servicios vamos a centrarnos en:

- /classes: esta sección permite clasificar el texto que se recibe como parámetro. Devuelve los tópicos a los que más se asemeja, un id : 0 significaría que el tópico es de relevancia alta para el texto introducido, un id : 1 relevancia media, y por último id : 2 relevancia baja, por debajo de ese nivel no se consideran el resto de los tópicos.

```
[  
  {  
    "description": "string",  
    "id": 0,  
    "name": "string"  
  }  
]
```

Ilustración 27 Imagen que muestra el esquema de la respuesta JSON del servicio /classes.

- /inferences este método sigue la misma idea de /classes, como clasifica-

dor de textos, también devuelve la lista de tópicos con los que está relacionado, siguiendo la misma idea que hemos descrito, siendo id : 0 el más relevante e id : 2 el menos relevante. Sin embargo, este método añade un parámetro a la respuesta de la API que no incluye /classes, se trata de un vector de números, de 0 a 1.

Este vector tiene un tamaño igual al número de tópicos, y muestra el grado de correlación que tiene el texto introducido con cada tópico, cuanto más se acerca a 1, más relevante es el tópico en el texto. La totalidad del vector suma a 1.

```
{
  "topics": [
    {
      "description": "string",
      "id": 0,
      "name": "string"
    }
  ],
  "vector": [
    0
  ]
}
```

Ilustración 28 Imagen que muestra la respuesta del servicio /inferences de la API de librAlry.

- /topics este método devuelve algunas características todos los tópicos del modelo. Estas características incluyen la lista de palabras más relevantes y el nombre del tópico.

```
[
  {
    "id": 0,
    "name": "00817",
    "description": "parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"
  },
  {
    "id": 1,
    "name": "0081",
    "description": "niño,familia,pobreza,situación,millón,pais,través,iniciativa,entidad,vulnerable"
  },
  {
    "id": 2,
    "name": "0083",
    "description": "sanitario,salud,nino,traves,paciente,iniciativa,eyder,mejor,situación,poner"
  },
  {
    "id": 3,
    "name": "0082",
    "description": "alimento,donación,familia,producto,traves,nino,iniciativa,situación,millon,solidario"
  },
  {
    "id": 4,
    "name": "0084",
    "description": "educación,nino,programa,formación,joven,educativo,traves,alumno,profesional,digital"
  },
  {
    "id": 5,
    "name": "0085",
    "description": "mujer,genero,igualdad,hombre,profesional,laboral,programa,parte,nino,femenino"
  },
  {
    "id": 6,
    "name": "0087",
    "description": "energético,energia,renewable,eléctrico,emisión,instalación,sostenible,millon,consumo,permitir"
  },
  {
    "id": 7,
    "name": "0088",
    "description": "trabajo,laboral,parte,director,organización,empleo,sector,empleo,programa,profesional"
  },
  {
    "id": 8,
    "name": "0089",
    "description": "innovación,sostenible,sector,tecnología,desarrollo,parte,traves,impacto,iniciativa,solución"
  },
  {
    "id": 9,
    "name": "00810",
    "description": "discapacidad,laboral,entidad,traves,programa,parte,inclusión,iniciativa,trabajo,sociedad"
  },
  {
    "id": 10,
    "name": "00811",
    "description": "sostenible,ciudad,vehículo,eléctrico,parte,movilidad,traves,cambio,sostenibilidad,iniciativa"
  },
  {
    "id": 11,
    "name": "00812",
    "description": "sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,traves"
  },
  {
    "id": 12,
    "name": "00813",
    "description": "climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millon"
  },
  {
    "id": 13,
    "name": "00816",
    "description": "organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"
  },
  {
    "id": 14,
    "name": "00815",
    "description": "natural,basuraleza,residuo,entorno,especie,naturaleza,diversidad,conservación,ave,ambiental"
  },
  {
    "id": 15,
    "name": "0089",
    "description": "agua,potable,millon,pais,hidrónico,recurso,sostenible,aceamiento,consumo,acceso"
  },
  {
    "id": 16,
    "name": "00814",
    "description": "plástico,marino,residuo,mar,playa,oceano,iniciativa,agua,conservación,parte"
  }
]
```

Ilustración 29 Imagen que muestra la respuesta del servicio /topics

- /topics/{id}: Opcionalmente esta extensión del método, permite pasar como parámetro un id, de manera que solo devuelve las características del tópico al que corresponde el id. Este método incluye también una puntuación de su entropía. La entropía es la media estadística de incertidumbre de una variable, desde un punto de vista del lenguaje, la entropía es la información producida de media por cada token en un lenguaje.²⁶

```
{
  "id": 1,
  "name": "0053",
  "description": "niño,familie,pobreza,situación,millón,pais,traves,iniciativa,entidad,vulnerable",
  "entropy": 0.18448644557488903
}
```

Ilustración 302 La imagen muestra la respuesta a la petición /topics/1.

- /topics/{id}/words: Por último, este método GET, permite conocer una lista más extensa de los tokens de cada tópico. Además, cada tópico se describe mediante una distribución de densidad sobre todo el vocabulario, es decir, la suma del score de todas las palabras para un tópico dado es igual a uno. Esos valores se van ajustando mediante maximización/minimización hasta conseguir que el modelo se ajuste con el corpus de entrenamiento. En la siguiente imagen podemos ver un ejemplo para el ODS 1:

```
[
  {
    "value": "ni\u00f1o",
    "score": 0.015663573678385972
  },
  {
    "value": "familia",
    "score": 0.007661530603558355
  },
  {
    "value": "pobreza",
    "score": 0.00723589001447178
  },
  {
    "value": "situaci\u00f3n",
    "score": 0.006959223631565506
  },
  {
    "value": "mill\u00f3n",
    "score": 0.006852813484293862
  },
  {
    "value": "pa\u00eds",
    "score": 0.006107942453392356
  },
  {
    "value": "trav\u00e9s",
    "score": 0.005256661275219205
  },
  {
    "value": "iniciativa",
    "score": 0.0048644557488903
  }
]
```

```
    "score": 0.004873584745041287
},
{
  "value": "entidad",
  "score": 0.004767174597769644
},
{
  "value": "vulnerable",
  "score": 0.004511790244317698
},
{
  "value": "parte",
  "score": 0.0043202519792287395
},
{
  "value": "programa",
  "score": 0.003937175449050821
},
{
  "value": "necesidad",
  "score": 0.003915893419596493
},
{
  "value": "ayuda",
  "score": 0.003894611390142164
},
{
  "value": "ayudar",
  "score": 0.003894611390142164
},
{
  "value": "vida",
  "score": 0.0038520473312335063
},
{
  "value": "crisis",
  "score": 0.0036392270366902187
},
{
  "value": "acción",
  "score": 0.003596662977781561
},
{
  "value": "económico",
  "score": 0.0034902528305099175
},
{
  "value": "mundo",
  "score": 0.0033838426832382734
}
```

Ilustración 31 La imagen muestra la respuesta a la petición /topics/1.

Tras recopilar la lista de las palabras de cada uno de los ODS en respuestas JSON como la mostrada en la imagen superior. Encontramos ciertas palabras como “través”, “parte”, estas palabras se repiten en muchos de los tópicos, y a priori no añaden información significativa, si no que serían mas bien palabras de unión.

También hay palabras como “niño” o “sostenibilidad” que aun que, si específicas de los ODS frente a textos sobre otros temas, no aportan mucha información para distinguir entre ODS, por estar, de nuevo, representados en muchos de los tópicos.

Decidimos continuar con el modelo tal cual, al no saber si estas palabras contienen algún grado de información relevante para el modelo.

6.5 Fase 4 Testeo del Modelo

Se ha desarrollado una web básica²⁷ que permita al usuario introducir un texto, y que tras realizar una llamada al método /classes de la API generada, devuelva en orden los Objetivos de desarrollo a los que pertenece.

Para ello se ha hecho uso de los lenguajes HTML y CSS para la parte visual de la página, y por otro lado de PHP para recibir el texto aportado por el usuario y hacer una llamada a un script en Python (a través de la función exec) para que sea este el que haga el procesado del texto y la llamada al servicio /classes. Después de esto PHP recoge el resultado y lo devuelve a la interfaz web.



Ilustración 32 Imagen que muestra la interfaz web desarrollada para el testeo del modelo.

La web se ha desplegado en local, haciendo uso del servidor web Apache de XAMPP, paquete de software que además de incluir sistema de gestión de bases de datos en MySQL, también tiene interprete de lenguaje PHP²⁸.

Bienvenido

Este proyecto pretende ayudar a Se basa en tecnologías de Inteligencia Artificial y Machine Learning para crear un modelo de tópicos y poder reconocer tokens representativos de cada ODS.

Poné el texto a clasificar en el cuadro inferior destinado a ello y pulsa enviar, el modelo de tópicos hará el resto y te devolverá una lista de tópicos ordenados de más a menos representativos.

La respuesta incluye una lista con todos los ODS que son representativos en el texto, el campo id marca el grado de relevancia, siendo el id=0 el mas relevante, incluye también una lista de las palabras o "tokens" más relevantes para cada ODS.

¡PRUEBEO!

Introducir texto

Enviar

```
[{"id":0,"name":"ODS7","description":"energético, energía renovable, eléctrico, emisión, instalación, sostenible, millón, consumo, permitir"}, {"id":1,"name":"ODS12","description":"sostenible, producto, plástico, sostenibilidad, residuo, compromiso, consumo, responsable, parte, través"}, {"id":2,"name":"ODS1","description":"niño, familia, pobreza, situación, millón, país, través, iniciativa, entidad, vulnerable"}]
```

Ilustración 33 Imagen que muestra la respuesta de la web al introducir el texto de Corresponsables titulado “ODS7. BBVA se suma a la Alianza Europea del Hidrógeno Limpio”

Para el correcto funcionamiento de esta interfaz web, ha de estar arrancado el Docker con la imagen generada por librAIry con el modelo de tópicos, para poder realizar peticiones a dicha API. También es necesario que la carpeta donde se encuentra el fichero PHP con la información de la página incluya los siguientes elementos:

-  CSS
-  images
-  import_requests
-  Web_ODS.php

Ilustración 343 Imagen que muestra la distribución de los documentos dentro de la carpeta htdocs de XAMPP.

Adicionalmente se creará un fichero .txt que con el nombre “texto” que servirá para método de comunicación entre el fichero PHP y el script de Python. Sin embargo, este fichero será creado por la función write() en caso de no existir por lo que no es necesario que esté inicialmente en la carpeta. En nuestro caso se encuentra en la carpeta htdocs dentro de XAMPP, para que podamos lanzar en local la web.

Se han dividido las pruebas en dos, unos casos de prueba con artículos de corresponsables.com que han sido añadidos recientemente y no formaron parte del corpus de entrenamiento.

La idea de estos dos casos de prueba es comparar que tan bien funciona el clasificador con artículos de noticias como los que usaron para su entrenamiento, frente a documentos que, aunque de los mismos temas, de ligeramente distinta estructura y probablemente diversidad de vocabulario. Previsiblemente, el clasificador debería funcionar mejor con los primeros.

Evaluación del modelo usando documentos etiquetados, primero con crónicas de la ONU extraídos de la página oficial de las naciones unidas. A continuación, se muestran los JSON devueltos por la petición al servicio /classes.

6.5.1 Pruebas con textos extraídos de la ONU:

Cada uno de los documentos utilizados en este apartado son parte de una secuencia de crónicas publicadas por parte de la ONU, y cada texto es específico de un ODS.

ODS 1. Fin de la pobreza

```
[{"id":0,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,páis,tráves,iniciativa,entidad,vulnerable"}, {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 35 respuesta JSON a la petición /classes con el texto Objetivo 1—Poner fin a la pobreza en todas sus formas y en todo el mundo

Como podemos observar en la imagen superior, el modelo identifica correctamente al Objetivo de poner fin a la pobreza como el más significativo en el texto, otorgándole un id=0.

ODS 2. Hambre cero

```
[{"id":0,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":1,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS2","description":"alimento,donación,familia,producto,tráves,nino,iniciativa,situación,millón,solidario"}]
```

Ilustración 36 respuesta JSON a la petición /classes con el texto Objetivo 2—Abrir nuevos caminos para la nutrición, la agricultura y los sistemas alimentarios²⁹

Esta imagen muestra como el modelo falla en reconocer al Objetivo de hambre cero como el más relevante del texto, sin embargo, sí que reconoce que la presencia de los tokens de este ODS en el texto, otorgándole una relevancia baja id=2.

ODS 3. Salud y bienestar

```
[{"id":0,"name":"ODS3","description":"sanitario,salud,niño,tráves,paciente,iniciativa,ayudar,mayor,situación,poner"}, {"id":1,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS17","description":"parte,organización,director,tráves,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 37 respuesta JSON a la petición /classes con el texto Objetivo 3—Los Objetivos de Desarrollo Sostenible y un 2030 más saludable³⁰

En este caso el modelo funciona como se esperaba, dando mayor relevancia tópico que esperábamos, dándole al ODS una relevancia alta id=0.

ODS 4. Educación de calidad

```
[{"id":0,"name":"ODS17","description":"parte,organización,director,tráves,sostenible,iniciativa,sociedad,responsable,sector,mundo"}, {"id":1,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,tráves,alumno,profesional,digital"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 38 respuesta JSON a la petición /classes con el texto Objetivo 4—Educación en la Agenda para el Desarrollo Sostenible Despues de 2015³¹

La imagen mostrada muestra como el modelo falla en reconocer al Objetivo de Educación de calidad como el más predominante del texto, apareciendo el ODS 17 en su lugar. A pesar de esto el modelo sí que identifica como relevancia media el ODS 4, otorgándole un id=1.

ODS 5. Igualdad de género

Objetivo 5—Lograr la igualdad de género y empoderar a todas las mujeres y las niñas: ¿se deja algo en el tintero el ODS 5?³²

```
[{"id":0,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":1,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}]
```

Ilustración 39 respuesta JSON a la petición /classes con el texto Objetivo 5—Lograr la igualdad de género y empoderar a todas las mujeres y las niñas: ¿se deja algo en el tintero el ODS 5?³³

En esta prueba el modelo de tópicos reconoce exitosamente al Objetivo de “Igualdad de género” como el más relevante para el texto de prueba por lo que le otorga un id=0.

ODS 6. Agua limpia y saneamiento

```
[{"id":0,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":1,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":2,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}]
```

Ilustración 40 respuesta JSON a la petición /classes con el texto Objetivo 6—Hacer frente al reto: posibilitar el acceso al agua limpia y potable en todo el mundo³⁴

En esta imagen superior apreciamos que el id=0 está asociado al “ODS 6”, esto supone un acierto por parte del clasificador, dado que el texto que se le pasó como parámetro era uno de ese tópico.

ODS 7. Energía asequible y no contaminante

```
[{"id":0,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":1,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}]
```

Ilustración 41 respuesta JSON a la petición /classes con el texto Objetivo 7—Garantizar el acceso a una energía asequible, fiable, sostenible y moderna para todos³⁵

Esta prueba no es del todo exitosa, dado que a pesar de reconocer tokens pertenecientes al Objetivo de “Energía asequible y no contaminante”, solo le atribuye una relación parcial dándole un id=1 y lo confunde con el ODS 12 de “Producción y consumo responsable”.

ODS 8. Trabajo decente y crecimiento económico

```
[{"id":0,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,nño,femenino"} {"id":2,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 42 respuesta JSON a la petición /classes con el texto Objetivo 8—Análisis del Objetivo 8 relativo al trabajo decente para todos³⁶

La prueba representada en esta imagen muestra como el modelo ha sido incapaz de identificar una presencia relevante del Tópico 8 de “Trabajo decente y crecimiento económico”, no asignándole así ningún valor al Id, y no mostrándose en el JSON devuelto. Por contraste muestra como tópico más relevante el ODS 16 de “Paz, justicia e instituciones sólidas”.

ODS 9. Industria, innovación e infraestructura

```
[{"id":0,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":1,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 43 respuesta JSON a la petición /classes con el texto Objetivo 9—Infraestructuras con un futuro sostenible³⁷

Está imagen al igual que la mostrada en la prueba VIII no muestra presencia del tópico del texto de prueba en el JSON de respuesta del modelo de tópicos. Identifica en cambio como tópico predominante al ODS 6 “Agua limpia y saneamiento”.

ODS 10. Reducción de las desigualdades

```
[{"id":0,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,nño,femenino"} {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 44 respuesta JSON a la petición /classes con el texto Objetivo 10—Por qué es importante combatir la desigualdad³⁸

En esta prueba el modelo de tópicos muestra como tópico más relevante al ODS 16 “Paz, justicia e instituciones sólidas”, no considerando el de Reducción de desigualdades, que era el tema del texto de prueba, ninguna relevancia considerable, ni siquiera una baja.

ODS 11. Ciudades y comunidades sostenibles

```
[{"id":0,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"} {"id":1,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":2,"name":"ODS3","description":"sanitario,salud,niño,tráves,paciente,iniciativa,ayudar,mayor,situación,poner"} {"id":2,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,tráves,iniciativa,entidad,vulnerable"}]
```

Ilustración 45 respuesta JSON a la petición /classes con el texto Objetivo 11—Las ciudades desempeñarán un papel importante en la consecución de los Objetivos de Desarrollo Sostenible³⁹

Al igual que la prueba anterior, en este caso el clasificador falla en reconocer el tópico en cuestión como relevante, no asignando un valor de id al ODS 11.

ODS 12. Producción y consumo responsables

```
[{"id":0,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"} {"id":1,"name":"ODS10","description":"discapacidad,laboral,entidad,tráves,programa,parte,inclusión,iniciativa,trabajo,sociedad"} {"id":2,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,tráves,impacto,iniciativa,solución"} {"id":2,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"} {"id":2,"name":"ODS7","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":2,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,tráves,alumno,profesional,digital"} {"id":2,"name":"ODS2","description":"alimento,donación,familia,producto,tráves,niño,iniciativa,ayudar,mayor,situación,poner"} {"id":2,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,oceano,iniciativa,agua,conservación,parte"} {"id":2,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":2,"name":"ODS15","description":"natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"} {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,tráves"} {"id":2,"name":"ODS11","description":"sostenible,ciudad,vehículo,eléctrico,parte,movilidad,tráves,cambio,sostenibilidad,iniciativa"} {"id":2,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,tráves,iniciativa,entidad,vulnerable"} {"id":2,"name":"ODS17","description":"parte,organización,director,tráves,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 48 respuesta JSON a la petición /classes con el texto Objetivo 12—Garantizar modalidades de consumo y producción sostenibles: un requisito esencial para el desarrollo sostenible⁴⁰

En esta prueba ha de resaltarse que todos los tópicos parecen tener algún grado de representación en el texto, aunque la mayoría bajas. En lo relativo al tópico que se estaba probando, este aparece con un id=2, por lo que se le atribuye una baja relevancia.

ODS 13. Acción por el clima

```
[{"id":0,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"} {"id":1,"name":"ODS7","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":1,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":2,"name":"ODS10","description":"discapacidad,laboral,entidad,tráves,programa,parte,inclusión,iniciativa,trabajo,sociedad"} {"id":2,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,tráves,impacto,iniciativa,solución"} {"id":2,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"} {"id":2,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,tráves,alumno,profesional,digital"} {"id":2,"name":"ODS2","description":"alimento,donación,familia,producto,tráves,niño,iniciativa,situación,millón,solidario"} {"id":2,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,oceano,iniciativa,agua,conservación,parte"} {"id":2,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":2,"name":"ODS15","description":"natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"} {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,tráves"} {"id":2,"name":"ODS11","description":"sostenible,ciudad,vehículo,eléctrico,parte,movilidad,tráves,cambio,sostenibilidad,iniciativa"} {"id":2,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,tráves,iniciativa,entidad,vulnerable"} {"id":2,"name":"ODS17","description":"parte,organización,director,tráves,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 49 respuesta JSON a la petición /classes con el texto Objetivo 13—La adopción de medidas urgentes para combatir el cambio climático – los ODS y el acuerdo de París sobre el clima⁴¹

Al igual que en la prueba anterior, en este texto el modelo de tópicos adjudica una relevancia por lo menos baja a todos los tópicos, sin embargo, en esta prueba el nivel de representación asignado al tópico de prueba es un grado superior, asignándole un id=1, es decir una relevancia media.

ODS 14. Vida submarina

```
[{"id":0,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}, {"id":1,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,océano,iniciativa,agua,conservación,parte"}, {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"}]
```

Ilustración 50 respuesta JSON a la petición /classes con el texto Objetivo 14—Conservar y utilizar sosteniblemente los océanos, los mares y los recursos marinos para el desarrollo sostenible⁴²

Para la prueba del ODS 14, que nos habla de la vida submarina, el modelo de tópicos asigna mayor relevancia a otro tópico, el del ODS 13 “Acción por el clima”, sin embargo, si logra identificar la relevancia que tiene el tópico de vida submarina en el texto, dándole un id=1.

ODS 15. Vida de ecosistemas terrestres

```
[{"id":0,"name":"ODS15","description":"natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"}, {"id":1,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 51 respuesta JSON a la petición /classes con el texto Objetivo 15—Los árboles nos dejan ver el bosque: aprovechar al máximo las sinergias para lograr los Objetivos de Desarrollo Sostenible en un entorno restringido⁴³

Esta imagen muestra cómo se asigna la mayor relevancia al tópico probado, dándole un id=0 al ODS 15 “Vida de ecosistemas terrestres” que era lo esperado.

ODS 16. Paz, justicia e instituciones sólidas

```
[{"id":0,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}, {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":2,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,través,iniciativa,entidad,vulnerable"}]
```

Ilustración 52 respuesta JSON a la petición /classes con el texto Objetivo 16—La paz en el marco para después de 2015: aprobación, aplicación y supervisión⁴⁴.

En esta prueba el tópico de “Paz, justicia e instituciones sólidas” no aparece en el JSON de respuesta, no considerándolo representativo del texto, ni siquiera en baja medida.

ODS 17. Alianzas para lograr los objetivos

```
[{"id":0,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":0,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}, {"id":1,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"}, {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 53 respuesta JSON a la petición /classes con el texto Objetivo 17—Posibilitar un futuro sostenible con la acción conjunta de países y comunidades: una Alianza Mundial para el Desarrollo Sostenible revitalizada⁴⁵

La imagen muestra la última prueba realizada sobre textos de las naciones unidas, en este caso el modelo ha identificado correctamente al ODS 17 como relevante para el texto, aunque solo le ha asignado una relevancia media, identificando el ODS 5 de “Igualdad de género” como tópico más influyente.

Resumen pruebas.

Hemos obtenido un total de 5 tópicos que han sido identificados como los más relevantes dentro de la prueba de su prueba, 7 tópicos que han sido reconocidos como relevantes, pero no como los más relevantes, con valores en el campo id de 1 o 2 y por último 5 tópicos que no han sido reconocido como relevantes en los textos marcados con su ODS.

Estas pruebas muestran que el modelo de tópicos funciona, ya que de lo contrario la tasa de acertar el tópico más relevante sería aproximadamente de

$\frac{1}{17}$ frente al $\frac{5}{17}$ que muestra el modelo, sin embargo, su tasa de acierto no es todo lo preciso que cabría esperar. Aun que somos conscientes que haría falta más pruebas para tener resultados más estadísticamente relevantes.

En el apartado de “Conclusiones” analizaremos algunas de las posibles razones por las que el modelo no ha obtenido mejores resultados.

6.5.2 Pruebas con textos extraídos de la web de Corresponsables:

Pruebas sobre textos extraídos de Corresponsables.com, se han utilizado los últimos artículos añadidos a la página, a fecha de 25/05/2021. Todas las imágenes muestran la respuesta en JSON del modelo de tópicos, al hacer una petición POST a la API de libraIry, usando como parámetro a un artículo en formato texto etiquetado con su correspondiente tópico.

ODS 1. Fin de la pobreza

```
[{"id":0,"name":"ODS2","description":"alimento,donación,familia,producto,través,niño,iniciativa,situación,millón,solidario"} {"id":1,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"} {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"}]
```

Ilustración 54 respuesta JSON a la petición del servicio /classes con un texto del ODS 1 de la plataforma Corresponsables.

El texto del artículo en cuestión trata de la donación de productos de limpieza a familias en riesgo de pobreza⁴⁶. En la imagen se puede observar como el clasificador no ha atribuido una relevancia significativa al tópico de fin de la pobreza, no otorgándole siquiera un id.

ODS 2. Hambre cero

```
[{"id":0,"name":"ODS2","description":"alimento,donación,familia,producto,través,niño,iniciativa,situación,millón,solidario"} {"id":1,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,través,iniciativa,entidad,vulnerable"} {"id":2,"name":"ODS10","description":"desapacidad,laboral,entidad,través,programa,parte,inclusión,iniciativa,trabajo,sociedad"} {"id":2,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,iniciativa,solución"} {"id":2,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"} {"id":2,"name":"ODS7","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":2,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"} {"id":2,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,través,alumno,profesional,digital"} {"id":2,"name":"ODS3","description":"sanitario,salud,niño,través,paciente,iniciativa,ayudar,mayor,situación,poner"} {"id":2,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,oceano,iniciativa,agua,conservación,parte"} {"id":2,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":2,"name":"ODS15","description":"natural,basuraleza,residue,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"} {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":2,"name":"ODS11","description":"sostenible,ciudad,vehículo,eléctrico,parte,movilidad,través,cambio,sostenibilidad,iniciativa"} {"id":2,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 55 respuesta JSON a la petición del servicio /classes con un texto del ODS 2 de la plataforma Corresponsables.

La noticia en cuestión cuenta la operación de recogida de alimentos por parte de un gran supermercado en España⁴⁷. El modelo identifica correctamente que el tema tópico dominante en este texto es el de Hambre cero, que coincide con la etiqueta asignada al texto. Ha de resaltarse que, según el modelo de tópicos, todos los ODS tienen presencia en el texto.

ODS 3. Salud y bienestar

```
[{"id":0,"name":"ODS3","description":"sanitario,salud,niño,través,paciente,iniciativa,ayudar,mayor,situación,poner"}, {"id":1,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"}, {"id":1,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,responsable,sector,mundo"}, {"id":2,"name":"ODS18","description":"discapacidad,laboral,entidad,través,programa,parte,inclusión,iniciativa,trabajo,sociedad"}, {"id":2,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,iniciativa,solución"}, {"id":2,"name":"ODS10","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":2,"name":"ODS55","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}, {"id":2,"name":"ODS4","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS22","description":"alimento,donación,familia,producto,través,niño,iniciativa,situación,millón,solidario"}, {"id":2,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,océano,iniciativa,agua,conservación,parte"}, {"id":2,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"}, {"id":2,"name":"ODS15","description":"natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"}, {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}, {"id":2,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"}, {"id":2,"name":"ODS11","description":"sostenible,ciudad,vehículo,electrónico,parte,movilidad,través,cambio,sostenibilidad,iniciativa"}, {"id":2,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,través,iniciativa,entidad,vulnerable"}]
```

Ilustración 56 respuesta JSON a la petición del servicio /classes con un texto del ODS 3 de la plataforma Corresponsables.

El articulo trata sobre la importancia de los ensayos clínicos, especialmente en tiempos de pandemia⁴⁸. Como muestra la respuesta JSON, el clasificador identifica correctamente el tópico de “Salud y bienestar” como el predominante en el texto, asignándole un id=0 es decir, relevancia alta así que la predicción concuerda con la etiqueta del texto. Ha de resaltarse que el modelo asigna relevancia por lo menos baja a todos los tópicos del modelo.

ODS 4. Educación de calidad

```
[{"id":0,"name":"ODS1","description":"niño,familia,pobreza,situación,millón,pais,través,iniciativa,entidad,vulnerable"}, {"id":1,"name":"ODS6","description":"agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"}, {"id":2,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,través,alumno,profesional,digital"}, {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}]
```

Ilustración 57 respuesta JSON a la petición del servicio /classes con un texto del ODS 4 de la plataforma Corresponsables.

El Articulo trata del absentismo escolar en países en desarrollo⁴⁹, la imagen de la respuesta JSON muestra como el modelo de tópicos encuentra representación en forma de tokens del tópico 4. Sin embargo, solo otorga una relevancia baja a este tópico, otorgando relevancias mas altas a los temas de “Fin de la pobreza” y “Agua limpia y saneamiento” en ese orden.

Aun que entendemos que en un texto clasificado como ODS 4, debería predominar dicho tema, no es de extrañar que identifique como temas relevantes los identificados por los ODS 1 y 6, al tratar el tema del agua potable, y países en desarrollo.

ODS 5. Igualdad de género

```
[{"id":0,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"}, {"id":1,"name":"ODS55","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":2,"name":"ODS3","description":"sanitario,salud,niño,través,paciente,iniciativa,ayudar,mayor,situación,poner"}]
```

Ilustración 58 respuesta JSON a la petición del servicio /classes con un texto del ODS 5 de la plataforma Corresponsables.

El articulo analizado por el clasificador trata del plan internación de women4It “mujeres para la informática” para reducir la brecha de género en el sector digital⁵⁰. Como muestra la imagen, el modelo identifica la relevancia del tópico 5 en el texto, aunque la categoriza como media y encuentra más similitudes con los tokens recogidos para el ODS 8.

ODS 6. Agua limpia y saneamiento

```
[{"id":0,"name":"ODS3","description":"sanitario,salud,niño,través,paciente,iniciativa,ayudar,mayor,situación,poner"}, {"id":1,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}, {"id":2,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}]
```

Ilustración 59 respuesta JSON a la petición del servicio /classes con un texto del ODS 6 de la plataforma Corresponsables.

El artículo es una entrevista al representante de Bluephage, empresa que realiza análisis microbiológico del agua a través de biotecnología⁵¹. El clasificador falla en reconocer como relevante el tópico en estudio “agua limpia y saneamiento” mientras encuentra más relevante el tópico de salud y bienestar.

ODS 7. Energía asequible y no contaminante

```
[{"id":0,"name":"ODS7","description":"energético, energía, renovable, eléctrico, emisión, instalación, sostenible, millón, consumo, permitir"} {"id":1,"name":"ODS10","description":"discapacidad, laboral, entidad, través, programa, parte, inclusión, iniciativa, trabajo, sociedad"} {"id":1,"name":"ODS9","description":"innovación, sostenible, sector, tecnología, desarrollo, parte, través, impacto, iniciativa, solución"} {"id":1,"name":"ODS8","description":"trabajo, laboral, parte, director, organización, empleado, sector, empleo, programa, profesional"} {"id":1,"name":"ODS5","description":"mujer, género, igualdad, hombre, profesional, laboral, programa, parte, niño, femenino"} {"id":1,"name":"ODS4","description":"educación, niño, programa, formación, joven, educativo, través, alumno, profesional, digital"} {"id":1,"name":"ODS2","description":"alimento, donación, familia, producto, través, niño, iniciativa, situación, millón, solidario"} {"id":1,"name":"ODS3","description":"sanitario, salud, niño, través, paciente, iniciativa, ayudar, mayor, situación, poner"} {"id":1,"name":"ODS14","description":"plástico, marino, residuo, mar, playa, océano, iniciativa, agua, conservación, parte"} {"id":1,"name":"ODS6","description":"agua, potable, millón, país, hídrico, recurso, sostenible, saneamiento, consumo, acceso"} {"id":1,"name":"ODS15","description":"natural, basuraleza, residuo, entorno, especie, naturaleza, biodiversidad, conservación, ave, ambiental"} {"id":1,"name":"ODS16","description":"organización, parte, director, sostenible, compromiso, sociedad, valor, gestión, sostenibilidad, responsable"} {"id":1,"name":"ODS13","description":"climático, emisión, cambio, sostenible, reducir, energía, carbono, parte, compromiso, millón"} {"id":1,"name":"ODS12","description":"sostenible, producto, plástico, sostenibilidad, residuo, compromiso, consumo, responsable, parte, través"} {"id":1,"name":"ODS11","description":"sostenible, ciudad, vehículo, eléctrico, parte, movilidad, través, cambio, sostenibilidad, iniciativa"} {"id":1,"name":"ODS1","description":"niño, familia, pobreza, situación, millón, país, través, iniciativa, entidad, vulnerable"} {"id":1,"name":"ODS17","description":"parte, organización, director, través, sostenible, iniciativa, sociedad, responsable, sector, mundo"}]
```

Ilustración 60 respuesta JSON a la petición del servicio /classes con un texto del ODS 7 de la plataforma Corresponsables.

Este documento de muestra trata sobre las perspectivas de nuevos combustibles no contaminantes⁵², tema a priori alineado con el tema de Energía asequible y no contaminante. El resultado de la clasificación muestra como el modelo a identificado correctamente al ODS 7 como el tópico más relevante para el texto. A destacar que en este texto aparecen representados todos los tópicos, esto ya lo hemos visto en alguna otra prueba, sin embargo, en este caso resulta impactante que todos tienen asignada una relevancia media, exceptuando el tópico de energía sostenible que tiene relevancia alta.

ODS 8. Trabajo decente y crecimiento económico

```
[{"id":0,"name":"ODS8","description":"trabajo, laboral, parte, director, organización, empleado, sector, empleo, programa, profesional"} {"id":1,"name":"ODS17","description":"parte, organización, director, través, sostenible, iniciativa, sociedad, responsable, sector, mundo"} {"id":2,"name":"ODS11","description":"sostenible, ciudad, vehículo, eléctrico, parte, movilidad, través, cambio, sostenibilidad, iniciativa"}]
```

Ilustración 61 respuesta JSON a la petición del servicio /classes con un texto del ODS 8 de la plataforma Corresponsables.

El articulo analizado por el modelo trata sobre diversas formulas para abordar una reconstrucción de la economía⁵³, este tema va alineado por lo menos con uno de los temas del tópico de “trabajo decente y crecimiento económico” que vamos a probar. Como vemos el clasificador a reconocido correctamente que el tópico principal del texto es el del ODS 8, asignándole un id=0.

ODS 9. Industria, innovación e infraestructura

```
[{"id":0,"name":"ODS9","description":"innovación, sostenible, sector, tecnología, desarrollo, parte, través, impacto, iniciativa, solución"} {"id":1,"name":"ODS17","description":"parte, organización, director, través, sostenible, iniciativa, sociedad, responsable, sector, mundo"} {"id":2,"name":"ODS5","description":"mujer, género, igualdad, hombre, profesional, laboral, programa, parte, niño, femenino"}]
```

Ilustración 62 respuesta JSON a la petición del servicio /classes con un texto del ODS 9 de la plataforma Corresponsables.

El articulo nos habla de la innovación en la química para hacer los procesos en la industria más eficientes⁵⁴, a priori parece que encaja bastante bien con el ODS 9. Y así lo certifica el modelo de tópicos, que le atribuye una relevancia alta al tópico esperado.

ODS 10. Reducción de las desigualdades

```
[{"id":0,"name":"ODS10","description":"discapacidad,laboral,entidad,través,programa,parte,inclusión,initiativa,trabajo,sociedad"} {"id":1,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,initiativa,solución"} {"id":1,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"} {"id":1,"name":"ODS7","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,nño,femenino"} {"id":1,"name":"ODS4","description":"educación,nño,programa,formación,joven,educativo,través,alumno,profesional,digital"} {"id":1,"name":"ODS2","description":"alimento,donación,familia,producto,través,nño,initiativa,situación,millón,solidario"} {"id":1,"name":"ODS3","description":"sanitario,salud,nño,través,paciente,initiativa,ayudar,mayor,situación,poner"} {"id":1,"name":"ODS14","description": "plástico,marino,residuo,mar,playa,oceano,initiativa,agua,conservación,parte"} {"id":1,"name":"ODS6","description": "agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":1,"name":"ODS15","description": "natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"} {"id":1,"name":"ODS16","description": "organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":1,"name":"ODS13","description": "climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":1,"name":"ODS12","description": "sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":1,"name":"ODS11","description": "sostenible,ciudad,vehículo,eléctrico,parte,movilidad,través,cambio,sostenibilidad,initiativa"} {"id":1,"name":"ODS1","description": "nño,familia,pobreza,situación,millón,pais,través,initiativa,entidad,vulnerable"} {"id":1,"name":"ODS17","description": "parte,organización,director,través,sostenible,initiativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 63 respuesta JSON a la petición del servicio /classes con un texto del ODS 10 de la plataforma Corresponsables.

El documento analizado para este tópico trata sobre la equidad e inclusión para garantizar igualdad de oportunidades entre toda la población⁵⁵. Vemos que en este caso también parece tener estrecha relación con el tópico analizado.

La imagen muestra como el clasificador correctamente detecta el ODS 10 como tópico principal tratado en el texto.

ODS 11. Ciudades y comunidades sostenibles

```
[{"id":0,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":1,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":2,"name":"ODS2","description":"alimento,donación,familia,producto,través,nño,initiativa,situación,millón,solidario"}]
```

Ilustración 64 respuesta JSON a la petición del servicio /classes con un texto del ODS 11 de la plataforma Corresponsables.

Para este tópico el artículo trata sobre el modelo de Cabify para simplificar la movilidad de las empresas y como Cabify compensa las emisiones generadas en estos viajes⁵⁶. Aunque el tema pueda tener relación con las ciudades y la movilidad más o menos sostenible dentro de ella, a priori no creemos que este texto no personifica la esencia del ODS 11, quizás esto explica por qué el clasificador no encuentra relevante este tópico en el texto, no asignando un id al tópico en cuestión.

ODS 12. Producción y consumo responsables

```
[{"id":0,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"} {"id":1,"name":"ODS10","description":"discapacidad,laboral,entidad,través,programa,parte,inclusión,initiativa,trabajo,sociedad"} {"id":1,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,initiativa,solución"} {"id":1,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"} {"id":1,"name":"ODS7","description":"energético,energía,renewable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"} {"id":1,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,nño,femenino"} {"id":1,"name":"ODS4","description":"educación,nño,programa,formación,joven,educativo,través,alumno,profesional,digital"} {"id":1,"name":"ODS2","description":"alimento,donación,familia,producto,través,nño,initiativa,situación,millón,solidario"} {"id":1,"name":"ODS3","description":"sanitario,salud,nño,través,paciente,initiativa,ayudar,mayor,situación,poner"} {"id":1,"name":"ODS14","description": "plástico,marino,residuo,mar,playa,oceano,initiativa,agua,conservación,parte"} {"id":1,"name":"ODS6","description": "agua,potable,millón,pais,hídrico,recurso,sostenible,saneamiento,consumo,acceso"} {"id":1,"name":"ODS15","description": "natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"} {"id":1,"name":"ODS16","description": "organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"} {"id":1,"name":"ODS13","description": "climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"} {"id":1,"name":"ODS11","description": "sostenible,ciudad,vehículo,eléctrico,parte,movilidad,través,cambio,sostenibilidad,initiativa"} {"id":1,"name":"ODS1","description": "nño,familia,pobreza,situación,millón,pais,través,initiativa,entidad,vulnerable"} {"id":1,"name":"ODS17","description": "parte,organización,director,través,sostenible,initiativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 65 respuesta JSON a la petición del servicio /classes con un texto del ODS 12 de la plataforma Corresponsables.

La noticia nos habla de nuevos avances sostenibles en la producción de productos de limpieza, los cuales son muy contaminantes⁵⁷. El resultado del modelo muestra como este es capaz de identificar al ODS 12 como el principal del texto.

ODS 13. Acción por el clima

```
[{"id":0,"name":"ODS13","description":"climático,emisión,cambio,sostenible,reducir,energía,carbono,parte,compromiso,millón"}, {"id":1,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}, {"id":1,"name":"ODS12","description":"sostenible,producto,plástico,sostenibilidad,residuo,compromiso,consumo,responsable,parte,través"}, {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}]
```

Ilustración 66 respuesta JSON a la petición del servicio /classes con un texto del ODS 13 de la plataforma Corresponsables.

El tema tratado en el texto es un proyecto para reducir el impacto ambiental de la hostelería en Canarias⁵⁸. La imagen muestra como el modelo detecta mayoritariamente tokens del ODS 13 y le asigna la mayor relevancia en el texto.

ODS 14. Vida submarina

```
[{"id":0,"name":"ODS7","description":"energético,energía,renovable,eléctrico,emisión,instalación,sostenible,millón,consumo,permitir"}, {"id":1,"name":"ODS17","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,iniciativa,solución"}, {"id":2,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 67 respuesta JSON a la petición del servicio /classes con un texto del ODS 14 de la plataforma Corresponsables.

El modelo falla en reconocer como relevante el tópico de “vida submarina” como relevante en el texto, esto es sorprendente ya que el texto habla de un seminario sobre la pesca sostenible⁵⁹, tema muy ligado al tópico en cuestión.

ODS 15. Vida de ecosistemas terrestres

```
[{"id":0,"name":"ODS15","description":"natural,basuraleza,residuo,entorno,especie,naturaleza,biodiversidad,conservación,ave,ambiental"}, {"id":1,"name":"ODS14","description":"plástico,marino,residuo,mar,playa,oceano,iniciativa,agua,conservación,parte"}, {"id":2,"name":"ODS9","description":"innovación,sostenible,sector,tecnología,desarrollo,parte,través,impacto,iniciativa,solución"}]
```

Ilustración 68 respuesta JSON a la petición del servicio /classes con un texto del ODS 15 de la plataforma Corresponsables.

El modelo nos indica que el tópico más relevante en el texto es el de “Vida de ecosistemas terrestres”, lo cual concuerda con la etiqueta del texto utilizado, texto que trata los residuos en el medio natural y la basuraleza⁶⁰.

ODS 16. Paz, justicia e instituciones sólidas

```
[{"id":0,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}, {"id":1,"name":"ODS8","description":"trabajo,laboral,parte,director,organización,empleado,sector,empleo,programa,profesional"}, {"id":2,"name":"ODS5","description":"mujer,género,igualdad,hombre,profesional,laboral,programa,parte,niño,femenino"}, {"id":2,"name":"ODS16","description":"organización,parte,director,sostenible,compromiso,sociedad,valor,gestión,sostenibilidad,responsable"}]
```

Ilustración 69 respuesta JSON a la petición del servicio /classes con un texto del ODS 16 de la plataforma Corresponsables.

La imagen nos muestra la respuesta JSON del modelo tras recibir un texto que trata sobre la importancia de la casilla solidaria para construir una sociedad más justa⁶¹, se puede ver como el modelo no considera predominante el tema de “Paz, justicia e instituciones sólidas” en el texto, otorgándole tan solo una relevancia baja. Quizá esto se deba por la doble acepción de la palabra “Justicia”, la cual en el texto habla de justicia social e inclusión, tema más relevante en otros tópicos, mientras que la justicia referida en el ODS 16, refiere a la justicia jurídica. Aun que puede que este sea un mal texto para probar la eficacia del clasificador, se ha mantenido por tres razones:

Era el último texto disponible, y la prueba debía realizarse sin manipular el texto escogido, ya que esto podría llevar a adulterar los resultados.

Nos muestra que, aun que un texto haya sido etiquetado con un ODS concreto por los profesionales de corresponsables, el texto puede estar mal etiquetado. Los textos de prueba también contienen este tipo de textos “mal” etiquetados, así que el modelo debería ser capaz de identificarlos.

A pesar de todo esto, el modelo logra identificar la presencia del tópico en cuestión en el artículo, aunque solo sea una relevancia baja.

ODS 17. Alianzas para lograr los objetivos

```
[{"id":0,"name":"ODS10","description":"discapacidad,laboral,entidad,través,programa,parte,inclusión,iniciativa,trabajo,sociedad"}, {"id":1,"name":"ODS4","description":"educación,niño,programa,formación,joven,educativo,través,alumno,profesional,digital"}, {"id":2,"name":"ODS2","description":"alimento,donación,familia,producto,través,niño,iniciativa,situación,millón,solidario"}, {"id":2,"name":"ODS17","description":"parte,organización,director,través,sostenible,iniciativa,sociedad,responsable,sector,mundo"}]
```

Ilustración 70 respuesta JSON a la petición del servicio /classes con un texto del ODS 17 de la plataforma Corresponsables.

La última prueba realizada corresponde a un texto que trata sobre proyectos de atención a la infancia, de Hero en colaboración con la ONG Cruz roja⁶². La respuesta muestra como el modelo identifica la presencia del tópico de “alianzas para lograr los objetivos” aunque sea de manera leve, también identifica la presencia del ODS 4 y 10 como más relevantes, lo cual es lógico viendo el tema del que trata el artículo.

Resumen de resultados:

Muestra 9 aciertos en asignar al tópico correspondiente la mayor relevancia en el texto, frente a 4 textos identificados como relevantes, aunque no los más relevantes y 4 no importantes, a pesar de tratarse de artículos etiquetados con ese tópico. Es una tasa de aciertos mayor que la de las pruebas con textos obtenidos de la ONU. A pesar de esto y de que más de la mitad de los tópicos han sido correctamente identificados como los más relevantes, consideramos que los resultados podrían ser mejores y comentaremos en el apartado “futuros pasos” como llegar a ellos.

Como hemos indicado con anterioridad, el número de pruebas es muy pequeño para sacar conclusiones certeras, esto se debe principalmente a la dificultad de obtener textos, circunstancia que ha limitado tanto la fase de entrenamiento como la de pruebas.

7 Conclusiones:

Es indudable que, al haber sido entrenado con documentos de una única fuente, ha inferido patrones que, siendo comunes en los artículos de corresponsables de cada tópico, no tienen por qué necesariamente ser representativos del tópico en cuestión, razón por la cual se explica una mayor tasa de acierto en los textos extraídos de la misma fuente que la de los textos de entrenamiento.

Una de las conclusiones que puede sacarse es que los ODS comparten múltiples similitudes, como muestran muchos tokens en común, esto podría hacernos ver que los tópicos no son tan independientes como se esperaría. Esto puede explicar algunos de los resultados parcialmente correctos, en los que se identifica el tópico, pero se le atribuye un peso menor del que a priori tenemos previsto que tenga. Una posible solución para esto sería eliminar las palabras que hemos identificado como comunes a muchos ODS, “sostenibilidad” o “niño”, aunque esto podría dificultar a que el modelo detectase aquellos textos ajenos a los ODS.

Viendo con retrospectiva, la elección de la fuente puede que no fuese la adecuada, a pesar de contar con muchos de los requisitos que habíamos propuesto, por lo menos, debería haberse complementado con una mayor diversidad de textos de múltiples fuentes, evitando así sobreentrenar el modelo. Hubiese sido preferible coger una fuente oficial como la ONU, pero tras haber explorado múltiples alternativas, no nos fue posible de encontrar una con una suficiente cantidad de textos, incluso se contactó con empleados de la ONU y del SDG PATHFINDER para solicitar colaboración.

Otro aspecto clave por el que decimos que, la elección de la fuente puede que no haya sido la adecuada, es el número de textos de cada uno de los ODS, mientras que contamos con un número total de textos superior a 10.000, solo habiendo 47 documentos del ODS 15, 77 documentos del ODS 14 y 112 documentos del ODS 6. Frente a 1503 documentos del ODS 3, 1257 documentos del ODS 8 y 1249 documentos del ODS 17.

Durante el proceso de recopilación se eliminaron una cantidad de artículos no determinada de la página de corresponsables, se intentó hacer uso de la página WayBack machine, que guarda versiones antiguas de las páginas web, sin embargo, la velocidad de navegación es muy lenta lo que hace imposible automatizar el proceso de extracción, por lo que nos vimos obligados a continuar con la última versión disponible en la web actual, con menos documentos de lo esperado.

Otro error fue confiar plenamente en el correcto etiquetado de los textos, aunque no hubiese manera automática de verificar este etiquetado, de esta, o de ninguna de las demás fuentes que se contemplaron en el proyecto, solo los artículos de la ONU con etiquetado, como los textos usados de prueba, podrían ser considerados como “gold standards”, sin embargo, no hay hemos encontrado una fuente con estas características.

Eliminar palabras de conexión en la fase de extracción de librAlry, como “través” (de la frase “a través de”), añadiéndolas al documento en solr dedicado a detectar este tipo de palabras, de esta forma la tasa de acierto del modelo debería verse positivamente afectada.

El bot que hace el scraping de la página es lento, debido a los retardos introducidos para que no se produzcan errores. Además de esto, el ritmo al que se añaden textos nuevos a la página no es muy elevado, sobre todo para los

ODS que menos atención reciben por esta página (ODS 6 y ODS 14) haciendo que la escalabilidad limitada. A pesar de esto, se han encontrado en 2 meses unos 500 textos nuevos, de mantenerse actualizado, podría generar un modelo más real.

En futuras versiones, podría aprovecharse la ventaja de que la herramienta librAIry es capaz de procesar múltiples idiomas para hacer un entrenamiento “cross-lingual”, con textos por ejemplo en castellano y en inglés.

Se debería también construir un clasificador SVM o de redes Bayesianas y que en el recaiga la decisión de identificar si un texto está relacionado o no a un tópico, esto tendría como ventajas que se podrían filtrar aquellos textos que no pertenecen a ningún ODS, ya que, de momento, el modelador de tópicos simplemente asigna el tópico más relevante en el texto, aunque este texto no trate sobre ninguno de los tópicos. Otra ventaja sería que se podrían mitigar en cierto grado los “falsos negativos” producidos, ya que ayudaría a formar patrones entre distintos tópicos.

Una última idea que surgió en la fase final del proyecto consiste en pedir a Organizaciones no Gubernamentales específicas de cada ODS documentación, por ejemplo, a “médicos sin fronteras” documentos para entrenar el ODS de “salud y bienestar” y a “Greenpeace” textos sobre “Acción contra el clima”.

En términos generales el rendimiento del modelo de tópicos ha sido más bajo que el esperado, hemos centrado gran parte del tiempo en encontrar una gran cantidad de textos para la fase de entrenamiento, y creíamos que con eso obtendríamos buenos resultados cosa que finalmente no ha resultado del todo acertada.

8 Bibliografía

- [1] A. Requena Mesa, "BIG DATA: La evolución de los datos", OpenWebinars.net, 2021.[Online]. Available: <https://openwebinars.net/blog/big-data-la-evolucion-de-los-datos/>. [Accessed: 26-Feb- 2021].
- [2]"Corresponsables.com España", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/>. [Accessed: 10- Mar- 2021].
- [3] "Expenses by SDG | United Nations - CEB", Unsceb.org, 2021. [Online]. Available: <https://unsceb.org/expenses-sdg>. [Accessed: 14- Mar- 2021].
- [4] "BIG DATA: La evolución de los datos", OpenWebinars.net, 2021. [Online]. Available: <https://openwebinars.net/blog/big-data-la-evolucion-de-los-datos/>. [Accessed: 08- Mar - 2021].
- [5] Vercelli, A., & Marotias, A. (2007). Guía de licencias Creative Commons. Consultado en Enero, 26, 2011.
- [6] "Tipos de licencias para contenidos digitales - Puesta en marcha - Emprendedor -Cecarm", Cecarm.com, 2021. [Online]. Available: <https://www.cecarm.com/emprendedor/puesta-en-marcha/tipos-de-licencias-para-contenidos-digitales-32010>. [Accessed: 22- May- 2021].
- [7] Sánchez Ruiz, Alberto (2019). Clasificación inicial de proyectos de una plataforma de crowdfunding de acuerdo con los Objetivos de Desarrollo Sostenible. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S. de Ingenieros Informáticos (UPM), Madrid, España.
- [8] UNSD SDGs API", Unstats.un.org, 2021. [Online]. Available: <https://unstats.un.org/sdgapi/swagger/>. [Accessed: 27- Feb- 2021].
- [9]"United Nations Digital Library System", United Nations Digital Library System, 2021. [Online]. Available: <https://digitallibrary.un.org/>. [Accessed: 05- Mar- 2021].
- [10]"Actualidad", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad>. [Accessed: 10- Mar- 2021].
- [11] "SDG Pathfinder", Sdg-pathfinder.org, 2021. [Online]. Available: <https://sdg-pathfinder.org/sdg/no-poverty>. [Accessed: 01- Mar- 2021].
- [12] "Google Colaboratory", Colab.research.google.com, 2021. [Online]. Available: <https://colab.research.google.com/drive/1Qkw9HtyT9f1R-yjMqfYLF3pXGMBK4ARD>. [Accessed: 22- Mar- 2021].
- [13] "libraIry", Librairy.github.io, 2021. [Online]. Available: <http://librairy.github.io/>. [Accessed: 23- Mar- 2021].
- [14] Red Hat Containers. What is Docker? [Online]. Available:

<https://www.redhat.com/es/topics/containers/what-is-docker>. [Accessed: 31- Mar- 2021]

- [15] Universidad de Alcalá Master Data Analytics [online] Available at: <<https://www.master-data-scientist.com/uah/data-analytics/>> [Accessed 01-Mar- 2021].
- [16] Application Programming Interface (API) Research: A Review of the Past to Inform the Future Joshua Ofoeda (University of Professional Studies, Accra, Ghana), Richard Boateng (University of Ghana Business School, Accra, Ghana) and John Effah (University of Ghana Business School, Accra, Ghana)
- [17] Libraries.mit.edu. 2021. API para recursos académicos | Publicación académica - Bibliotecas del MIT.
- [18] Crummy.com. 2021. Beautiful Soup Documentation — BeautifulSoup 4.9.0 documentation. [online] Available at: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>> [Accessed 18- Mar- 2021].
- [19] Chromedriver.chromium.org. 2021. ChromeDriver - WebDriver for Chrome. [online] Available at: <<https://chromedriver.chromium.org/>> [Accessed 20 Mar 2021].
- [20] PyPI. 2021. selenium. [online] Available at: <<https://pypi.org/project/selenium/>> [Accessed 20 Mar 2021].
- [21] Instituto de Ingeniería del Conocimiento. 2021. Expertos en Big Data e Inteligencia Artificial - IIC. [online] Available at: <<http://www.iic.uam.es/>> [Accessed 1 May 2021].
- [22] E. Cambria and B. White, "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]," in IEEE Computational Intelligence Magazine, vol. 9, no. 2, pp. 48-57, May 2014, doi: 10.1109/MCI.2014.2307227. [Accessed 3 May 2021].
- [23] Calle Velasco, Guillermo de la (2014). Modelo basado en técnicas de procesamiento de lenguaje natural para extraer y anotar información de publicaciones científicas. Tesis (Doctoral), E.T.S. de Ingenieros Informáticos (UPM). Ieeexplore.ieee.org. 2021. Jumping NLP Curves: A Review of Natural Language
- [24] Blei, David M. et al. "Latent Dirichlet Allocation." J. Mach. Learn. Res. 3 (2003): 993-1022. Jmlr.org. 2021. [online] Available at: <<https://jmlr.org/papers/volume3/blei03a/blei03a.pdf>>[Accessed 20 May 2021].
- [25] GitHub. 2021. Julen-Perez/TFG_NLP_ODS. [online] Available at: <https://github.com/Julen-Perez/TFG_NLP_ODS/blob/main/tfg_bot_obtener_enlaces.py>
- [26] Blei, David M. et al. "Latent Dirichlet Allocation." J. Mach. Learn. Res. 3 (2003): 993-1022.
- Available at: <<http://www.cs.columbia.edu/~blei/papers/BleiNgJordan2003.pdf>> [Accessed 15 May 2021].
- [27] Perez/TFG_NLP_ODS";, GitHub, 2021. [Online]. Available: https://github.com/Julen-Perez/TFG_NLP_ODS/tree/main/web_ods.

-
- [28] Apache Friends [Online]. Available: <https://www.apachefriends.org/es/index.html>. [Accessed: 20 – May -2021].
 - [29] U. Nations, "Objetivo 1—Poner fin a la pobreza en todas sus formas y en todo el mundo | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-1-poner-fin-la-pobreza-en-todas-sus-formas-y-en-todo-el-mundo>. [Accessed: 21- May- 2021].
 - [30] U. Nations, "Objetivo 2—Abrir nuevos caminos para la nutrición, la agricultura y los sistemas alimentarios | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-2-abrir-nuevos-caminos-para-la-nutricion-la-agricultura-y-los-sistemas-alimentarios>. [Accessed: 21- May- 2021].
 - [31] U. Nations, "Objetivo 3—Los Objetivos de Desarrollo Sostenible y un 2030 más saludable | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-3-los-objetivos-de-desarrollo-sostenible-y-un-2030-mas-saludable>. [Accessed: 21- May- 2021].
 - [32] U. Nations, "Objetivo 4—Educación en la Agenda para el Desarrollo Sostenible Después de 2015 | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-4-educacion-en-la-agenda-para-el-desarrollo-sostenible-despues-de-2015>. [Accessed: 21- May- 2021].
 - [33] U. Nations, "Objetivo 5—Lograr la igualdad de género y empoderar a todas las mujeres y las niñas: ¿se deja algo en el tintero el ODS 5? | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-5-lograr-la-igualdad-de-genero-y-empoderar-todas-las-mujeres-y-las-ninas-se-deja-algo-en-el>. [Accessed: 21- May- 2021].
 - [34] U. Nations, "Objetivo 6—Hacer frente al reto: posibilitar el acceso al agua limpia y potable en todo el mundo | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-6-hacer-frente-al-reto-posibilitar-el-acceso-al-agua-limpia-y-potable-en-todo-el-mundo>. [Accessed: 21- May- 2021].
 - [35] U. Nations, "Objetivo 7—Garantizar el acceso a una energía asequible, fiable, sostenible y moderna para todos | Naciones Unidas", United Nations, 2021. [Online]. Available: [https://www.un.org/es/chronicle/article/objetivo-7-garantizar-el-acceso-una-energia-asequible-fiable-sostenible-y-moderna-para-todos#:~:text=El%20ODS%207%20\(%E2%80%9Cgarantizar%20el,que%20nos%20alcanza%20a%20tardos.&text=Las%20cuatro%20dimensiones%20del%20Objetivo,la%20sostenibilidad%20y%20la%20modernidad](https://www.un.org/es/chronicle/article/objetivo-7-garantizar-el-acceso-una-energia-asequible-fiable-sostenible-y-moderna-para-todos#:~:text=El%20ODS%207%20(%E2%80%9Cgarantizar%20el,que%20nos%20alcanza%20a%20tardos.&text=Las%20cuatro%20dimensiones%20del%20Objetivo,la%20sostenibilidad%20y%20la%20modernidad). [Accessed: 01- Jun- 2021].
 - [36] U. Nations, "Objetivo 8—Análisis del Objetivo 8 relativo al trabajo decente para todos | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-8-analisis-del-objetivo-8-relativo-al-trabajo-decente-para-todos>.

8-analisis-del-objetivo-8-relativo-al-trabajo-decente-para-to-
dos#:~:text=El%20Objetivo%208%20tiene%20como,el%20tra-
bajo%20decente%20para%20todos%E2%80%9D. [Accessed: 21- May-
2021].

- [37] U. Nations, "Objetivo 9—Infraestructuras con un futuro sostenible | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-9-infraestructuras-con-un-futuro-sostenible>. [Accessed: 21- May- 2021].
- [38] U. Nations, "Objetivo 10—Por qué es importante combatir la desigualdad | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-10-por-que-es-importante-combatir-la-desigualdad>. [Accessed: 21- May- 2021].
- [39] U. Nations, "Objetivo 11—Las ciudades desempeñarán un papel importante en la consecución de los Objetivos de Desarrollo Sostenible | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-11-las-ciudades-desempenaran-un-papel-importante-en-la-consecucion-de-los-objetivos-de#:~:text=M%C3%A1s%20de%20la%20mitad%20de,que%20contin%C3%BAa%20esa%20tendencia%20migratoria.&text=De%20ah%C3%AD%20que%20el%20Objetivo,sostenibles%E2%80%9D%C2%20sea%20tan%20importante>. [Accessed: 21- May- 2021]
- [40] U. Nations, "Objetivo 12—Garantizar modalidades de consumo y producción sostenibles: un requisito esencial para el desarrollo sostenible | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-12-garantizar-modalidades-de-consumo-y-produccion-sostenibles-un-requisito-esencial-para-el>. [Accessed: 21- May- 2021].
- [41] U. Nations, "Objetivo 13—La adopción de medidas urgentes para combatir el cambio climático – los ODS y el acuerdo de París sobre el clima | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-13-la-adopcion-de-medidas-urgentes-para-combatir-el-cambio-climatico-los-ods-y-el-acuerdo>. [Accessed: 21- May- 2021].
- [42] U. Nations, "Objetivo 14—Conservar y utilizar sosteniblemente los océanos, los mares y los recursos marinos para el desarrollo sostenible | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-14-conservar-y-utilizar-sosteniblemente-los-oceanos-los-mares-y-los-recursos-marinos-para>. [Accessed: 21- May- 2021].
- [43] U. Nations, "Objetivo 15—Los árboles nos dejan ver el bosque: aprovechar al máximo las sinergias para lograr los Objetivos de Desarrollo Sostenible en un entorno restringido | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-15-los-arboles-nos-dejan-ver-el-bosque-aprovechar-al-maximo-las-sinergias-para-lograr-los>. [Accessed: 21- May- 2021].
- [44] U. Nations, "Objetivo 16—La paz en el marco para después de 2015: aprobación, aplicación y supervisión | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-16-la-paz-en-el>

marco-para-despues-de-2015-aprobacion-aplicacion-y-supervision. [Accessed: 21- May- 2021].

- [45] U. Nations, "Objetivo 17—Posibilitar un futuro sostenible con la acción conjunta de países y comunidades: una Alianza Mundial para el Desarrollo Sostenible revitalizada | Naciones Unidas", United Nations, 2021. [Online]. Available: <https://www.un.org/es/chronicle/article/objetivo-17-posibilitar-un-futuro-sostenible-con-la-accion-conjunta-de-paises-y-comunidades-una>. [Accessed: 22- May - 2021].
- [46] "ODS1. Las fundaciones Cepsa y Persán donarán 90.000 kilos de productos de limpieza a 9.000 familias vulnerables", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods1-fundacion-cepsa-persan-productos-limpieza-familias-vulnerables> [Accessed: 22- May - 2021].
- [47] [3]"ODS2. Fesbal y Eroski comienzan una nueva edición de la "Operación kilo" de recogida de alimentos", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods2-fesbal-eroski-nueva-edicion-operacion-kilo-recogida-alimentos>. [Accessed: 22- May - 2021].
- [48] "ODS3. MSD pone en valor el papel crucial de los ensayos clínicos", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods3-msd-papel-crucial-ensayos-clinicos>. [Accessed: 22- May - 2021].
- [49] "ODS4. AUARA reduce el absentismo escolar de 48.000 niños de países en desarrollo gracias a la mejora de la salud y el agua potable", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods4-auara-absentismo-escolar-paises-en-desarrollo-salud-agua-potable>. [Accessed: 22- May - 2021].
- [50] "ODS5. Plan International apuesta con Women4IT por cerrar la brecha de género en el sector digital en la UE", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods5-plan-international-women4it-brecha-de-genero-sector-digital-ue>. [Accessed: 22- May - 2021].
- [51] "ODS6. Bluephage se convierte en la primera empresa biotecnológica especializada en el análisis microbiológico del agua en conseguir la certificación B Corp", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods6-bluephage-primer-a-empresa-biotecnologia-analisis-microbiologico-agua-certificacion-b-corp-entrevista-enric-queralt>. [Accessed: 22- May - 2021].
- [52] "ODS7. Nuevos combustibles, actualidad y perspectivas", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods7-nuevos-combustibles-actualidad-perspectivas-fundacion-corell>. [Accessed: 22- May - 2021].
- [53] "ODS8. NESI Global Forum 2021 buscará nuevas fórmulas disruptivas para reconstruir la economía", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods8-nesi-global-forum-2021-formulas-disruptivas-reconstruir-economia>. [Accessed: 22- May - 2021].
- [54] "ODS9. Cepsa impulsa la innovación en la química con su pionera

tecnología Detal", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods9-cepsa-innovacion-quimica-tecnologia-detal>. [Accessed: 22- May - 2021].

- [55] "ODS10. Liberty constituye el Consejo de Diversidad, Equidad e Inclusión para garantizar la igualdad de oportunidades", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods10-liberty-consejo-diversidad-equidad-inclusion-igualdad-oportunidades>. [Accessed: 22- May - 2021].
- [56] "ODS11. Cabify y Emburse Captio se unen para simplificar la movilidad de las empresas", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods11-cabify-emburse-captio-simplificar-movilidad-empresas>. [Accessed: 22- May - 2021].
- [57] "ODS12. Covestro crea un nuevo material para la limpieza basado en CO2", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods12-covestro-nuevo-material-limpieza-co2>. [Accessed: 22- May - 2021].
- [58] "ODS13. Veolia contribuye a la reducción del impacto ambiental de la industria hotelera en Canarias", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods13-veolia-reducir-impacto-ambiental-industria-hotelera-canarias>. [Accessed: 22- May - 2021].
- [59] "ODS 14. Jealsa promueve un seminario virtual sobre la pesca sostenible del futuro", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods14-jealsa-promueve-seminario-virtual-pesca-sostenible-futuro>. [Accessed: 22- May - 2021].
- [60] "ODS15. LIBERA presenta la app 'Basuraleza' para registrar los residuos abandonados en el medio natural", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/libera-app-basuraleza-residuos-medio-natural-seobirdlife-ecoembes>. [Accessed: 22- May - 2021].
- [61] "ODS16. El Tercer Sector explica cómo construir una sociedad más justa y conseguir ventajas al marcar la Casilla Empresa Solidaria", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods16-tercer-sector-justicia-ventajas-casilla-empresa-solidaria-asociacion-espanola-fundaciones>. [Accessed: 22- May - 2021].
- [62] "ODS17. Los voluntarios de Hero España participan en dos proyectos de Cruz Roja para atención a la infancia", Corresponsables.com España, 2021. [Online]. Available: <https://www.corresponsables.com/actualidad/ods17-voluntarios-hero-espana-dos-proyectos-cruz-roja-atencion-infancia>. [Accessed: 22- May - 2021].