

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

# Interface

# USER'S GUIDE

\* \* \* \* \* \* \* This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

November 2020

http://pharmproject.usal.es

This report constitutes Deliverable 19, for Work Package 4 of the PHARM project.

October 2020

© 2020 – PHARM, Preventing Hate Against Refugees and Migrants, – GA number 875217.

General contact: pharm.project.eu@gmail.com

For more information type the e-mail address of the corresponding author

Information may be quoted provided the source is stated accurately and clearly. This publication is also available via <a href="https://pharmproject.usal.es/">https://pharmproject.usal.es/</a>

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Copyright © Pharm Consortium, 2020

All rights reserved. No part of the report may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, and may not be quoted or cited, without prior permission in writing from the project coordinator.

The views expressed during the execution of the project, in whatever form and or by whatever medium, are the sole responsibility of the authors. The European Union is not liable for any use that may be made of the information contained therein.



# VERSION CONTROL SHEET

Deliverable number + title	D10. Guide of use of the interface	
Prepared by	AUTH TEAM	
Work package number	2	
Work package leader	Andreas Veglis	
Dissemination level (PU, CO)	PU	
Delivery date	30/11/2020	
Submission date to EC	30/11/2020	
Main authors	Lazaros Vrysis / Nikolaos Vryzas	
Reviewers	Carlos Arcila Calderón / Patricia Sánchez Holgado	

#### **REVISION HISTORY**

Version	Date	Summary of changes	Initials	Changes marked
1				

Para la interfaz de Internet del proyecto PHARM

Español version 1.0

# Índice

1	Info	orma	ción técnica	2
	1.1	Inte	erfaz de Internet	2
	1.2	Info	prmación técnica	3
	1.2	2.1	Scraping	3
	1.2	2.2	Estimación de fecha y hora	4
	1.2	2.3	Estimación de Geolocalización	4
	1.2	2.4	Detección de lenguaje	4
	1.2	2.5	Detección de discurso de odio	4
	1.2	2.6	Análisis de sentimiento	5
	1.3	Usu	arios	5
	1.4	Acc	iones	5
	1.4	4.1	Búsqueda (Search)	5
	1.4	4.2	Analizar (Analyze)	6
	1.4	4.3	Scrape	6
	1.4	1.4	Anotar (Annotate)	7
	1.4	4.5	Enviar (Submit)	8
	1.5	For	mato de Datos	9
2	Gui	ía de	la Interfaz 1	2
	2.1	Intr	oducción 1	2
	2.2	Hor	ne1	2
	2.3	Bus	car (Search)1	3
	2.4	Ana	ılizar (Analyze) 1	4
	2.5	Anc	otar (Annotate) 1	6
	2.6	Env	iar (Submit)1	6
	2.7	Aut	enticación o Login1	8
3	Gui	ía de	Anotación 1	9
	3.1	Intr	oducción 1	9
	3.2	Apr	obar 2	1

# 1 Información técnica

PHARM es un proyecto europeo financiado por la Unión Europea, dentro del programa *Derechos, Igualdad y Ciudadanía REC-RRAC-RACI-AG-2019* (GA n. 875217). El objetivo principal de **Preventing Hate Against Refugees and Migrantes** (PHARM) es monitorear y modelar el discurso de odio contra refugiados y migrantes en Grecia, Italia y España para predecir y combatir los delitos de odio y también contrarrestar sus efectos utilizando técnicas de vanguardia, como periodismo de datos y persuasión narrativa. El resultado principal será la identificación y reducción del discurso de odio en línea y la predicción de posibles delitos de odio.

# 1.1 Interfaz de Internet

PHARM presenta una interfaz web con un formato integrado semiestructurado orientado a almacenar y consultar noticias de múltiples fuentes y contenidos de redes sociales. La interfaz consta de una base de datos y una interfaz web front-end para exponer datos y funcionalidad a los usuarios. La primera pantalla que se presenta al visitante es la pantalla de inicio, que se presenta en la Figura 1. Además de brindar información básica sobre el proyecto, esta pantalla proporciona un punto de partida para acceder a los análisis soportados en un texto.



Figura 1 Pantalla principal de la interfaz de internet PHARM

Las principales acciones que admite la interfaz son:

• Búsqueda: la búsqueda se refiere a la consulta de la base de datos para filtrar y descargar datos de incitación al odio.

• Analizar: el análisis proporciona resultados de varios tipos de análisis de texto, como detección de incitación al odio, análisis de sentimientos y estimación de geolocalización.

• Scrape: esta acción permite extraer datos de texto de las plataformas sociales de Twitter y YouTube.

• Anotar: habilita la anotación de los datos recopilados.

• Enviar: esta funcionalidad permite agregar nuevos registros de incitación al odio a la base de datos.

Las funciones de análisis y búsqueda se exponen libremente a todos los usuarios que utilizan la interfaz, mientras que la autenticación es necesaria para acceder a las acciones Scrape, Annotate y Submit. Si los usuarios quieren ser parte del proyecto para contribuir con datos adicionales, deben contactar a nuestro equipo para solicitar dicho rol visitando la página de contacto de la interfaz.

# 1.2 Información técnica

Para el desarrollo de la interfaz se ha utilizado el framework web Python Flask. La elección está justificada, ya que el *procesamiento del lenguaje natural o natural language processing* (NLP) y los procedimientos de análisis de datos también están escritos en *Python*. Con este enfoque, toda la funcionalidad de la interfaz se puede incluir dentro de un proyecto común. El front-end de la interfaz gráfica de usuario se ha diseñado principalmente en Bootstrap, una biblioteca popular de *HTML, CSS y JavaScript*. Se han agregado bloques adicionales de HTML, CSS y JavaScript cuando es necesario. El proyecto *Flask* se ha implementado en una máquina virtual y se sirve mediante *Waitress*, un servidor WSGI de Python puro de calidad de producción con un rendimiento muy aceptable.

Se han implementado varias bibliotecas para implementar los métodos de análisis compatibles, junto con algoritmos personalizados que se han desarrollado específicamente para el proyecto PHARM. Algunos de los métodos más importantes permiten la extracción de datos, la estimación de la fecha y la hora, la estimación de la geolocalización, la detección del idioma y la detección del discurso del odio. Las funciones adicionales, como el análisis de opiniones, estarán disponibles en la interfaz una vez finalizada, de acuerdo con el cronograma del proyecto. Las principales opciones de funcionalidad y diseño de estos métodos se presentan a continuación. Para obtener más información técnica, visite el repositorio: PHARM Github Repository.

# 1.2.1 Scraping

El proyecto admite la recopilación de texto de *Twitter* a través de la *API* adecuada y el método de transmisión. *Tweepy* se utiliza para acceder a la *API de Twitter*. Se han desarrollado cuatro diccionarios para filtrar tweets que incluyen palabras clave en griego, inglés, español e italiano. La recopilación de comentarios de YouTube es compatible a través de la *API de Google*. Se requiere una consulta de búsqueda relevante para el tema (por ejemplo, "migración de refugiados") del usuario y se recopilan los comentarios de los resultados principales (videos).

# 1.2.2 Estimación de fecha y hora

Se ha implementado un método para detectar y estandarizar la información de fecha y hora a partir de metadatos y texto. Además de la ubicación y el idioma, cuando hay metadatos disponibles, PHARM utiliza información adicional relevante para el análisis del discurso de odio. Parte de esta información adicional, como la fecha o la hora, puede estar disponible en la mayoría de los casos en varios formatos diferentes. Esto introduce la necesidad de estandarización. Para satisfacer las necesidades de este requisito, se explotan los paquetes de Python *dateparser, datefinder y parsedatetime*, clasificados desde una mayor precisión hasta una mayor probabilidad de obtener un resultado. Si el método más preciso no detecta un objeto de fecha y hora, se llama al siguiente servicio. La detección se basa en el análisis de metadatos, donde la información de fecha suele estar presente. Si la detección de fecha y hora falla para todos los servicios de los metadatos, se aplica el mismo flujo de trabajo a los datos de texto.

# 1.2.3 Estimación de Geolocalización

Se ha implementado un método para detectar la geolocalización a partir de datos de texto. Para ello se ha utilizado la biblioteca de *geopy*, junto con el geocodificador *nominatim*. Las entidades nombradas (características lingüísticas) se aíslan de los textos, de acuerdo con el siguiente ranking: GPE (países, ciudades, estados), LOC (montañas, cuerpos de agua), FAC (edificios, aeropuertos, carreteras, etc.), ORG (empresas, agencias, instituciones, etc.) y se geocodifican.

# 1.2.4 Detección de lenguaje

PHARM procesa principalmente texto producido en los idiomas griego, italiano y español, pero muchas de las fuentes pueden tener contenido en otros idiomas extranjeros o dialectos locales. Para trabajar con estos tres idiomas nacionales, se ha codificado un procedimiento para detectar el idioma del texto mediático cuando no está debidamente declarado. Se adopta un enfoque recursivo para mejorar la solidez, aprovechando las bibliotecas *textblob, googletrans* y *langdetect.* Si falla un servicio, se solicita el resultado del siguiente.

# 1.2.5 Detección de discurso de odio

Se han implementado un par de métodos para encontrar términos clave para la detección del discurso del odio. Estos incluyen coincidencia simple de cadenas de texto, coincidencia aproximada de cadenas de texto con el uso de las métricas adecuadas, como distancia de Levenshtein, distancia de Damerau-Levenshtein, distancia de Jaro, distancia de Jaro-Winkler, comparación de aproximación de calificación de iguales, distancia de Hamming.

La coincidencia de términos también tiene como objetivo ser agnóstica de sufijos, acomodando los diversos sufijos que pueden existir en los sustantivos para muchos idiomas (es decir, el idioma griego presenta diferentes sufijos para el género o versiones en singular / plural). También se ha desarrollado un enfoque de vector de palabras, teniendo en cuenta el significado semántico de los términos. Se ha implementado un enfoque híbrido basado en un diccionario con frases predefinidas, junto con combinaciones de términos dinámicos (es decir, adjetivos combinados con sustantivos) y se está evaluando. Se carga un modelo de

lenguaje (según el idioma del texto), se siguen prácticas habituales (minúsculas, lematización, eliminación de palabras vacías y puntuación) y se buscan los términos específicos en el texto. Si se encuentra un término (o una combinación de términos), el texto se marca como "discurso de odio". También se está desarrollando un algoritmo de clasificación supervisada adicional que se integrará en la interfaz.

# 1.2.6 Análisis de sentimiento

Se están desarrollando dos métodos para el análisis de sentimientos y se integrarán en la interfaz, de acuerdo con el cronograma del proyecto. Estos incluyen un enfoque sin supervisión basado en el léxico y un enfoque supervisado que explota la tecnología de aprendizaje automático.

# 1.3 Usuarios

The interface specifies two types of users: the visitor and the contributor. A visitor can access the analysis and search modules, while the contributor has full access to all available actions. The basic work-flows for a visitor (unauthenticated user) and a contributor (authenticated user) are defined as follows:

La interfaz especifica dos tipos de usuarios: el visitante y el colaborador. Un visitante puede acceder a los módulos de análisis y búsqueda, mientras que el colaborador tiene acceso completo a todas las acciones disponibles. Los flujos de trabajo básicos para un visitante (usuario no autenticado) y un colaborador (usuario autenticado) se definen a continuación:

#### **Colaborador (Contributor)**

- Buscar (Search) -> Analizar (Analyze)
- Scrape -> Anotar (Annotate) -> Enviar (Submit)

#### Visitante (Visitor)

• Buscar (Search) -> Analizar (Analyze)

# 1.4 Acciones

Las acciones principales que admite la interfaz son cinco: Buscar, Analizar, Extraer, Anotar y Enviar. A continuación se ofrece una descripción más detallada de estas funciones.

# 1.4.1 Búsqueda (Search)

Una de las principales funcionalidades de la interfaz es navegar por los registros de discurso de odio (textos) de la base de datos. El usuario puede ver todos los resultados o aplicar una variedad de filtros (por ejemplo, fuente, idioma, fecha). En detalle, los filtros disponibles son:

- Selección de fuente (Twitter, Facebook, artículos web, comentarios web).
- Selección de idioma (inglés, griego, italiano, español).
- Selección de fecha y hora (mostrar los resultados dentro de un período específico).

• Filtro de palabras clave (una consulta de búsqueda para encontrar coincidencias en textos).

El usuario puede obtener una vista previa de los resultados, elegir la vista "Simple" o "Científica", descargarlos como un archivo CSV o JSON, o mostrar información más detallada para cada registro.

# 1.4.2 Analizar (Analyze)

Cuando se selecciona un registro (o se coloca un texto en la pantalla de inicio), aparece una vista que presenta información detallada (Figura 2). La ubicación se marca en un mapa y los resultados de varios algoritmos de análisis de texto se presentan con gráficos (iconos, barras, etc.). Los resultados se refieren a la detección del discurso del odio (tanto para los métodos de clasificación supervisados como no supervisados), el análisis de sentimientos (para los métodos de clasificación supervisados y no supervisados), la detección frecuente de palabras y la recopilación de entidades del discurso del odio.

HARM Interface	Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
Analysis Enter text for hate speech and sentii	iment analysis	
"RT @GerryConnolly: Bill Barr has s	sold his soul to the president and sullied the reputation of the Department of Justice. He is a disgrace t"	æ
Perform Analysis Results		
Unsupervised Hate Speech D	Detection	
Supervised Hate Speech Dete	tection	
Unsupervised Sentiment Ana	alysis	
Supervised Sentiment Analys	sis	

Figura 2 Pantalla de análisis de la interfaz de internet

# 1.4.3 Scrape

Este módulo permite la recopilación masiva de datos de texto de dos plataformas populares: Twitter y YouTube. Un usuario puede recopilar datos de incitación al odio de Twitter, simplemente eligiendo el idioma deseado (griego, inglés, italiano o español) y lanzando el proceso presionando el botón "Scrape". Se inicia la transmision de datos y los tweets se recopilan en función de diccionarios específicos del idioma, desarrollados en el contexto del proyecto. El proceso se detiene después de un intervalo de tiempo configurable por el usuario. Se proporciona un enlace para descargar un archivo JSON que contiene los datos. Estos datos se pueden utilizar para cualquier tarea de procesamiento del lenguaje natural o *Natural Language Processing (NLP)*. El usuario puede repetir el proceso varias veces.

En el caso de YouTube, en lugar de seleccionar el idioma, se debe establecer una consulta de búsqueda. La consulta de búsqueda puede incluir términos de búsqueda individuales o una combinación de ellos, separados por una coma. El proceso es instantáneo en este caso y una vista en pantalla muestra los datos correspondientes. Estos datos se pueden descargar como un archivo CSV o JSON. Para obtener más información sobre los diccionarios de filtrado de Twitter y el formato de datos basado en JSON, puede consultar el repositorio: <u>PHARM</u> <u>GitHub repository</u>.

PHARM Interface	Home Search Submit	Scrape Annotate Instructions Contact About	Welcome admin Logout
Scrape Scrape comments from social	nedia		
Youtube Scraper Twitter Scraper Youtube scraper Enter search query			
Max results			
Scrape			

Figura 3 Pantalla de Scrape de la Interfaz de internet.

# 1.4.4 Anotar (Annotate)

El proceso de anotación está impulsado por la herramienta doccano. Doccano es un sistema de gestión de anotaciones para datos de texto y se puede utilizar para desarrollar conjuntos de datos para facilitar la clasificación, el análisis de opiniones, el etiquetado de entidades o las tareas de traducción. En el contexto del Proyecto PHARM, se utiliza para la clasificación de texto y cada registro debe etiquetarse con etiquetas específicas.

PHARM Interface	Home Search Submit Scrape <b>Annotate</b> Instructions Contact About	Welcome admin Logout
Annotate Visit our doccano installation	to annotate the data	
loccano 🛞	Live Demo 🗸 GitHub Login	
Search document	1/6	
About 6 results	Person p Loc I Org o Event e Date d Other z	
<ul> <li>Barack Hussein Obama II (born August 4, 1961) is an American</li> </ul>	Develo Unersia Obarra II.O. (kers. August 4, 4064 a. Vision	
The White House is the official residence and workplace of t The Democratic Party is one of the	American  attorney and politician who served as the 44th President of	
two major contemporary po	the United States • from January 20, 2009 • , to January 20, 2017 •	
Stanford Junior Unive	. A member of the <b>Democratic Party ×</b> , he was the first	
Donald John Trump (born June 14, 1946) is the 45th and curre	African American • to serve as president. He was previously a	
Silicon Valley (abbreviated as SV) is a region in the southe	United States Senator • from Illinois • and a member of the Illinois State Senate • .	
	< Prev Next >	

Figure 4 The Annotate Screen of the Internet Interface

# 1.4.5 Enviar (Submit)

Para agregar un nuevo registro (o un lote de registros), el usuario debe seleccionar "Submit" (Enviar) en el menú superior (Figura 5). La entrada de datos se puede ejecutar una a una o de forma masiva. Con respecto al primer método, el usuario debe configurar todos los datos (texto) y metadatos (fuente, idioma, fecha, odio, sentimiento, etc.) a través de los formularios de entrada correspondientes (es decir, campos de texto, botones de radio, etc.). Si los datos ya están formados adecuadamente (consulte el párrafo 1.5), también se pueden importar como un archivo JSON.

# 1.5 Formato de Datos

Teniendo en cuenta los requisitos del proyecto (es decir, el uso de información adicional relevante para el análisis del discurso de odio), las fuentes que se utilizan para recopilar el contenido relevante (es decir, artículos y comentarios del sitio web, comentarios de YouTube y tweets de Twitter), interoperabilidad y Consideraciones de compatibilidad para importar / exportar datos a aplicaciones de terceros que se explotan (es decir, la plataforma Doccano para anotaciones), se han establecido algunas especificaciones generales para el formato de datos.

M Interface	Home Search <b>Submit</b> Scrape Anno	tate Instructions Contact Abou	ut	Welcome admin Logout
Submit Submit a single or a batch of re	ecords to the PHARM database			
Single Text Batch File				
Submit single hate-speech related text	to the PHARM database			
Enter text here				
Insert tags (Optional)				
Choose text language:	с	hoose date (Optional):		
English	~	mm / dd / yyyy		
Choose text source: O Twitter O Yo	JTube O Facebook Comment O Faceboo	ok Post O Website Article O Webs	site Comment	]
Submit				

Figura 5 Vista de envío de la interfaz de Internet

El campo principal es el texto (contenido), acompañado por la identificación, anotaciones y metacampos (id, annotations, meta fields). El metacampo es un contenedor que incluye todos los metadatos adicionales. Se utiliza un conjunto fundamental/mínimo de metadatos para todas las plataformas (es decir, type, plang, pdate, phate, psent, pterms, ploc). Es más probable que estos campos se encuentren para todos los registros de diferentes fuentes. La siguiente tabla ofrece una descripción general del esquema de datos propuesto (Tabla 1).

Tabla 1 Los campos comunes del formato de datos PHARM

Campo	Descripción
id	Identificador único
annotations	Anotaciones de discurso de odio y sentimientos
meta	metadata
text	contenido
type	Tipo de texto (tweet, articulo, comentario, etc.)
plang	Detección de idioma via análisis de PHARM
pdate	Estimación de fecha via análisis de PHARM
phate	Estimación de discurso de odio via análisis de PHARM

psent	sentiment analysis via PHARM analysis
pterms	frequent terms collection via PHARM analysis
ploc	geolocation estimation via PHARM analysis

En los casos de web scraping, los metadatos dependen de los metadatos disponibles proporcionados por cada sitio y de la estructura específica del sitio, mientras que para los comentarios de YouTube y los tweets de Twitter, donde se utilizan las API correspondientes, se han seleccionado y recopilado metadatos específicos junto con los comentario de texto (Tabla 2)

Twitter	YouTube
tweet_id	comment_id
is_retweet	reply_count
is_quote	like_count
user_id	video_id
username	video_title
scr_name	channel
location	video_description
followers	author_id
friends	author_name
date	date

Tabla 2 Campos adicionales (metadatos) para los registros de YouTube y Twitter

Para una presentación más descriptiva, a continuación se muestran datos de ejemplo de las diferentes fuentes de datos.

## Facebook

{"id": "1", "annotations": [{"label": "hate", "user": 2}, {"label': "negative", 'user': 2}], "meta": {"id": "80056833", "type": "facebook\_comment", "source": "https://www.facebook.com/provinciale.rivista", "plang": "it", "pdate": "1571-10-15 00:00:00", "phate": "(nero cinesi)", "pterms": "", "ploc": "Italy, United States of America"}, "text": "andrebbe anche fatta con Venezia (principale flotta a Lepanto). 1571: le mie potentissime galeazze spaccheranno il culo all infedele ottomano. 2020: per favore basta quarantena, devo vendere lasagne sur-gelate in nero ai cinesi a 50 euro"}

#### Twitter

{"id": "2", "annotations": [{"label": "hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"type": "twitter\_comment", "date": "10/13/2020", "tweet\_id": 1315981643111432192, "is\_retweet": true, "is\_quote": false, "user\_id": 1025701121749340160, "username": "Christina Dim", "scr\_name": "ChristinaDim31", "location": "Αθήνα ", "followers": 3178, "friends": 4756, "quoted\_text": "", "pid": 27454420, "plang": "el", "pdate": "", "phate": "(Τουρκία μετανάστες)", "pterms": "", "ploc": "Turkey"}, "text": "RT @kanekos69: Εν τω μεταξύ αν γίνει στραβή με Τουρκία βλέπω μετανάστες εθελοντές στο μέτωπο και δεξιούς προσφυγες στα Παρίσια"}

#### YouTube

{"id": "3", "annotations": [{"label": "no\_hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"type": "youtube\_comment", "comment\_id": "Ugy-SPKz3HGo4OohJnfR4AaABAg", "reply\_count": 3, "like\_count": 30, "video\_id": "Gaz6UvRW0G8", "channel": "Mηδέν Ένα Μηδέν 010", "video\_itile": "0 1 0 ~ Πρόσφυγες", "video\_desc": "Lyrics/Raps - 0 1 0 Beat by Apo (Aισθήσεις) Recorded @ Blackspot Studio Mix/Master by Sativa Cover by SpyOne (Baseline Co.) 0 1 0 IG: ...", "author\_id": "UCPP-OugMmE8pNbWRbFv4YCA"}, "au-thor\_name": "GATE21QNZ", "rating": "none", "date": "2020-09-03T21:16:31Z", "plang": "el", "pdate": "2020-10-25 00:00:00", "phate": "(Αφγανιστάν ισλαμιστές)", "pterms": "", "ploc": "Afghanistan"}, "text": "Είμαι 25 ετών, ονομάζομαι Μήτσος ο μαλάκας, Μαζεύω ισλαμιστές και λιποτάκτες από μια τούρκο βάρκα, Το παίζω ανοιχτόκαρδος με τις τσέπες του μπαμπάκαα, Και όταν ο Μήτσος πήγε να κάνει φίλους στο Αφγανιστάν, Τον σφάξαν σαν τους προγόνους του, για χάρη του Ισλαμ."}

#### Artículo de Website

{"id": "4", "annotations": [{"label": "no\_hate", "user": 2}, {"label': "positive", 'user': 2}], "meta": {"id": "92860318", "type": "article", "source": "http://defencereview.gr/gnorizontas-ta-gallika-ploia-oi-fremm-kai-o/", "meta": "Άμυνα Ελλάδα 9 Σεπτεμβρίου 2020 18:22 ", "title": "Γνωρίζοντας τα γαλλικά πλοία: Οι FREMM και οι [email protected] (Video)", "lang": "el", "date": "2020-09-09 00:00:00", "hate": "", "terms": "", "loc": ""}, "text": "Συχνά πυκνά αναφερόμαστε στις γαλλικές ναυπηγικές σχεδιάσεις. Τα παρακάτω βίντεο που βρήκα-με είναι αντιπροσωπευτικά για τις δυνατότητες των πλοίων με πολύ καλά σκηνοθε-τημένα βίντεο και ενδιαφέροντα πλάνα. Αξίζει να τα δείτε: "}

#### *Comentario de Website*

{"id": "5", "annotations": [{"label": "no\_hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"id": "92860318", "type": "comment", "source": "http://defencereview.gr/gnorizontas-ta-gallika-ploia-oi-fremm-kai-o/", "lang": "el", "date": "", "hate": "", "terms": "", "loc": ""}, "text": "Το μπαραζ επεκτεινεται. Ελπιζω αυριο τετοια ωρα πανω κατω να μην κλαιμε"}

También puede obtener más información sobre el formato de datos y los campos que se utilizan visitando el repositorio del proyecto en PHARM GitHub repository.

# 2 Guía de la Interfaz

# 2.1 Introducción

La Interfaz de Internet permite buscar, ver, analizar, agregar y extraer textos de incitación al odio para el proyecto PHARM, proporcionando una interfaz gráfica de usuario amigable. Se puede acceder a la interfaz visitando el siguiente hipervínculo: <u>pharm-interface.usal.es</u>.

# 2.2 Home

La pantalla de inicio sirve como punto de partida para realizar análisis de texto en un texto definido por el usuario. Escriba (o pegue) el texto deseado en el cuadro de texto y presione el botón "Perform Analysis" (Realizar análisis), como se muestra en la Figura 6.



Figura 6 Funcionalidad de la pantalla principal Home

Se abre una nueva pestaña en el navegador web y se presentan los resultados del análisis ( Figura 9. ). Para obtener más información, consulte la Sección 2.4.

La navegación a todas las subpáginas de la Interfaz de Internet está disponible mediante el menú superior.

# 2.3 Buscar (Search)

Una de las acciones principales de la interfaz es la navegación a través de los registros de incitación al odio de la base de datos.

Los usuarios pueden visitar la "Search Screen" (Pantalla de búsqueda) seleccionando "Search" (Buscar en el menú superior. Después de configurar los filtros deseados, a saber, fuente, idioma, rango de fechas y consulta de búsqueda, deben presionar el botón "Search" (Buscar) (Figura 7). Los filtros de fuente e idioma son obligatorios, mientras que establecer una fecha y una consulta de búsqueda es opcional.

Se carga la "Search Results Screen" (Pantalla de resultados de búsqueda) y se muestra una lista de los resultados correspondientes (Figura 8).

El usuario puede obtener una vista previa de los registros, elegir la vista "Simple" o "Scientific" (Científica), descargarlos como un archivo CSV o JSON, eliminar un registro o mostrar información detallada sobre él.

Cabe señalar que la acción de eliminación solo está disponible para usuarios autenticados.

RM Interface	Home <b>Search</b> Submit Scrape An	notate Instructions Contact About	Welcome admin Logout
Search Search for hate speech cont	ent using multiple filters		
Choose source:			
☑ Twitter □ YouTube □ Faceboo	k Post 🛛 Facebook Comment 🗌 Website /	Article 🗌 Website Comment 🗌 All Sources	
Choose language:			
Italian			~
Posted from:		until:	
11/02/2020	0	11/19/2020	0
Filter results with query:			
Enter search query			
Search			

Figura 7 Funcionalidad de la pantalla de búsqueda

PHARM Interface Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
Search	
View the results of your query or go back for a new search	
82 results found with the selected filters	
Simple View Scientific View Scientific View	w) w)
.@MoniquePariatEU (Διευθύντρια Μετανάστευσης Κομισιόν): Ικαντράτου του χειρισμούς του υπ. Ναυτιλίας και του Λιμενικού στο μεταναστευτικό https://t.co/7VrOQWdHbz https://t.co/fEb9iHEDur more	
Φ Υ Λ Α Κ Η. Χωρίς δικαιώματα χωρίς παροχές χωρίς τίποτα. Όλοι οι λαθρομετανάστες να μπουν φυλακή. more 👕	
RT @giorgosptk: Λαθρομετανάστες που πήραν άσυλο και έφυγαν από την κόλαση της Μόριας για Αθήνα επιστρέφουν και πάλι στη Λέσβο.Ρε μλκι φαντ more	ç
RT @stoxos: ΔΕΙΤΕ ΤΟΥΣ ΣΙΧΑΜΕΝΟΥΣ!!! Σάμος: Λαθρομετανάστες έγλειφαν τρόφιμα σε σούπερ μάρκετ και γελούσαν - Δείτε βίντεο https://t.co/2 more	
Είμαστε σπόροι, που θα λουλουδιασουν. Η ευωδία μας θα μυρίζει λιβάνι και θα διώξει τους σατανάδες. Το Έθνος των Ελλήνων δε χάθηκε κάτω από 400 χρόνια ζυγού. Δε θα χαθούμε από μία ξεφτιλισμενη παγκόσμια ελίτ. Είμαστε το μέλλον. more	
RT @hellene_bc: Είμαστε σπόροι, που θα λουλουδιασουν. Η ευωδία μας θα μυρίζει λιβάνι και θα διώξει τους σατανάδες. Το Έθνος των Ελλήνων δ. more 🍵	
Σάμος: Λαθρομετανάστες έγλειφαν τρόφιμα σε σούπερ μάρκετ και γελούσαν - Δείτε βίντεο https://t.co/FFLJJ52WxM more 🍵	

Figura 8 Funcionalidad de la pantalla de resultados de búsqueda

# 2.4 Analizar (Analyze)

Cuando se selecciona un registro (o se coloca un texto en la "Pantalla de inicio"/"Home Screen"), aparece una vista que presenta información detallada (Figura 9.). Se presentan los resultados de los distintos algoritmos de análisis de texto. Los algoritmos se refieren a la detección del discurso del odio, el análisis de sentimientos, la detección frecuente de palabras, la estimación de geolocalización. El usuario puede realizar un nuevo análisis de texto sin salir de la pantalla, editando el texto en el cuadro de texto y presionando el botón "Perform Analysis" (Realizar análisis). En la parte inferior de la página, hay atajos para sugerir una corrección a nuestro equipo o hacer referencia al repositorio PHARM GitHub para encontrar más información sobre los algoritmos de análisis de texto implementados en la Interfaz.

PHARM Interface Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
Analysis Enter text for hate speech and sentiment analysis	
".@MoniquePariatEU (Διευθύντρια Μετανάστευσης Κομισιόν): Ικανοποίηση για τους χειρισμούς του υπ. Ναυτιλίας και του Λιμενικού μεταναστευτικό https://t.co/7VrOQWdHbz https://t.co/fEb9iHEDun"	) στο
Perform Analysis Results	
Unsupervised Hate Speech Detection	
Supervised Hate Speech Detection	
Unsupervised Sentiment Analysis Supervised Sentiment Analysis	
Geographical estimation of hate speech references	
Business Consulting none, Lagina 572 00 Vew larger map Torections Toreprov Acpleter Passias Cafe Bar Consulting Passias Cafe Bar Consulting Passias Cafe Bar Consultant Passias Cafe B	Club 4x4 Club 4x4 Club 4x4 Constantinos Dorr es to Dor reprint Agol ELEXOYAH Truck repart shop t
Coogle Map deta #2000 T Do you disagree with the hate-speech and sentiment analysis evaluation? Consider contributing your opinion to help us improve. For more information about the analysis algorithms please visit our GitHub page.	Terms of Use Report a map error

Figura 9. Funcionalidad de la pantalla de análisisScrape

Esta pantalla proporciona una funcionalidad para permitir la recopilación masiva de datos de texto de dos plataformas sociales, Twitter y YouTube.

Cuando se selecciona YouTube Scraper, aparece la pantalla que se muestra en la Figura 10. El usuario debe escribir algunos términos de búsqueda para recopilar comentarios relevantes de YouTube. También se puede configurar el número de resultados. Finalmente, se debe presionar el botón "Scrape" para cargar los resultados, que se presentan como se muestra en la Figura 8, mientras que los datos también se pueden descargar en formato CSV o JSON.

En el caso del scraping de Twitter, la pantalla tiene una forma similar (Figura 11 Funcionalidad de scraping de Twitter). Debe seleccionarse un idioma, junto con el intervalo de tiempo en el que estará activa la operación de scraping. Cuando se presiona el botón "Scrape" se genera un hipervínculo para descargar los datos correspondientes después de que pasa el intervalo de tiempo predefinido.

PHARM Interface	Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
Scrape		
Scrape comments from social	nedia	
Youtube Scraper Twitter Scraper		
Youtube scraper		
refugees		
Max results		
100		•
Scrape		

#### Figura 10 Funcionalidad de scraping de YouTube

PHARM Interface	Home Search S	Submit <b>Scrape</b> Annotate	Instructions Contact	About	Welcome admin Logout
Scrape Scrape comments from social	media				
Youtube Scraper Twitter Scraper Choose language:					_
Italian Streaming time (minutes): 10					
Scrape					

Figura 11 Funcionalidad de scraping de Twitter

# 2.5 Anotar (Annotate)

Para anotar datos, seleccione "Annotate" (Anotar) en el menú superior para navegar a la plataforma de anotación doccano. Para obtener más información, consulte la Sección3.

# 2.6 Enviar (Submit)

Para agregar un nuevo registro, el usuario debe seleccionar "Enviar" en el menú superior. La entrada de datos se puede ejecutar uno por uno o de forma masiva.

Con respecto al primer método, el usuario debe configurar todos los datos (texto) y metadatos (fuente, idioma, fecha, odio, sentimiento, etc.) a través de los formularios de entrada correspondientes. Los datos se agregan a la base de datos presionando el botón "Submit" (Enviar) (Figure 12 Single text submission functionality).

Si los datos ya están formados adecuadamente (consulte el párrafo 1.5), se pueden importar como un archivo (JSON). Seleccione "Batch File" (Archivo por lotes), browse for the file (buscar el archive), y presione el botón "Submit" (Enviar) (Figure 12 Single text submission functionality).

Una vez completada la importación de datos, se presenta la pantalla de resumen de importación (Figura 14). Esta pantalla muestra información útil sobre el total de registros que se han procesado, el número de duplicados o registros que ya están almacenados en la base de datos.

Además, las entradas no válidas, que carecen de campos obligatorios, se detectan y se notifica al usuario en consecuencia.

RM Interface	ome Search <b>Submit</b> Scrape Annotate Instructions Contact About	/elcome admin Logout
Submit Submit a single or a batch of re Single Text Batch File	rds to the PHARM database	
Submit single hate-speech related text	the PHARM database	1
Enter text here Insert tags (Optional) Tefugees ×	đ	
Choose text language:	Choose date (Optional):	
English	✓ mm/dd/yyyy	
Choose text source: O Twitter O You Insert source url (Optional) Submit	be O Facebook Comment O Facebook Post O Website Article O Website Comment	

Figure 12 Single text submission functionality



Figura 13 Funcionalidad de envío de registros por lotes



Figura 14 Importar pantalla de resumen de la interfaz de Internet

# 2.7 Autenticación o Login

Cuando el usuario selecciona una página que requiere autenticación (es decir, buscar, raspar y anotar), aparece la pantalla de inicio de sesión (Figura 15 Pantalla de autenticación de la interfaz de Internet ), donde se debe ingresar el nombre de usuario y la contraseña para continuar.



Figura 15 Pantalla de autenticación de la interfaz de Internet

Para obtener una cuenta autenticada, visite la "Pantalla de contacto" y comuníquese con nuestro equipo.

# 3 Guía de Anotación

La interfaz de Internet utiliza la herramienta Doccano para respaldar las tareas de anotación necesarias.

# 3.1 Introducción

Doccano es un sistema de gestión de anotaciones para datos de texto. Se puede utilizar para desarrollar conjuntos de datos para clasificación, análisis de opiniones, etiquetado de entidades o traducción. En el contexto del *Proyecto PHARM*, se utiliza con fines de clasificación de texto. Por lo tanto, cada entrada debe etiquetarse con etiquetas.

En primer lugar, el usuario puede echar un vistazo a la demostración en línea para el análisis de sentimientos (también se puede seguir el mismo procedimiento para la clasificación): <u>http://doccano.herokuapp.com/demo/text-classification</u>. El usuario puede acceder a la instalación de doccano visitando <u>http://pharm-doccano.usal.es</u>. También se puede acceder a la herramienta desde la pantalla correspondiente de la interfaz de Internet de PHARM (Figure 4 The Annotate Screen of the Internet Interface).

En pocas palabras, el usuario / anotador debe asignar etiquetas ("Odio" o "No odio", "Positivo", "Neutral" o "Negativo") ("Hate" or "No Hate", "Positive", "Neutral" or "Negative") a cada entrada. Las anotaciones se almacenan automáticamente.

Inicie sesión con sus credenciales como anotador (Figura 16 Pantalla de autenticación o Login de Doccano).

🕞 doccano	Live Demo 🗸 Git	Hub Login
Please login to see this page. Username		
▲ ✓		
Password		
Remember me		
Login Not registered yet 2 Sign up		
Not registered your organize		
		~

Figura 16 Pantalla de autenticación o Login de Doccano

Elija el proyecto en el que se le ha pedido que colabore (Figura 17). A continuación, lea el texto (para la entrada actual), elija "Hate" o "No Hate" (si el texto tiene contenido contra refugiados o migrantes o no) y márquelo como "Positivo", "Neu-tral" o "Negativo" (según el sentimiento del texto). Es obligatorio elegir una etiqueta "azul" y una "púrpura" para cada entrada, ya que la anotación se realiza para dos esquemas de clasificación individuales simultáneamente (Figura 18). Las etiquetas "azules" son para la detección del discurso de odio, mientras que las etiquetas "violetas" son necesarias para el análisis de opiniones. Las entradas anotadas se resaltan con un icono de marca de verificación. Finalmente, elija "siguiente" para pasar a la siguiente entrada.



Figura 17 Pantalla de selección de proyectos de Doccano



Figura 18 Pantalla de anotación y funcionalidad de selección de etiquetas de Doccano

# 3.2 Aprobar

Inicie sesión en doccano y seleccione el proyecto adecuado (siga los mismos pasos descritos anteriormente). Apruebe una anotación con el botón que se muestra en Figura 19. Las entradas aprobadas se muestran con un icono de pulgar hacia arriba.



Figura 19 Pantalla de anotación y funcionalidad de aprobación del doccano

# PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

# ABOUT PHARM (2020-2022)

Preventing Hate against Refugees and Migrants (PHARM)

Migration to Europe has grown in the last years in scale and complexity. The so called 'refugee crisis' and the migratory pressure is particularly acute in southern EU countries as the main entrance to the EU.

The main goal of PHARM project is to monitor and model hate speech against refugees and migrants in Greece, Italy and Spain in order to predict and combat hate crime and also counter its effects using cutting-edge techniques, such as data journalism and narrative persuasion. The activities distributed in 5 coordinated work packages include:

(i) Implementation of a conceptual and methodological common framework for large-scale analysis and detection of hate speech; (ii) Implementation and evaluation of machine learning approaches to model and predict hate crimes against refugees and migrants based on hate speech features; (iii) Survey journalists to understand how they inform and raise awareness about hate speech and how they can help building and disseminating counter-narratives based in data-driven news pieces; (iv) Creation, evaluation and dissemination of counter-narrative fictional stories adapted to different characteristics of citizens using large-scale narrative persuasion.

# COORDINATOR

University of Salamanca, Faculty of Social Sciences (ES)

# PARTNERS

University of Milan, (IT) Aristotle University of Thessaloniki (GR)

> VISIT: http://pharmproject.usal.es CONTACT US: pharm.project.eu@gmail.com FOLLOW US:

> > @Pharm\_project

Pharm\_project

**Pharm Project** 

Pharm Project