INFORME

BUENAS PRÁCTICAS Y EXPERIENCIAS

en la UE para analizar el **discurso de odio en línea**

Discurso de odio, racismo y xenofobia: mecanismos de alerta y respuesta coordinada

(AL-RE-CO)

Just/2017/Action Grants /REC PROGRAM

WP 2: PROTOCOLO Y SISTEMA DE INDICADORES CONTRA EL RACISMO, LA XENOFOBIA Y EL DISCURSO DE ODIO

ALRECO

Cofinanciado por el Programa Derechos, Igualdad y Ciudadanía de la Unión Europea

GOBIERNO DE ESPAÑA
MINISTERIO DE TRABAJO, MIGRACIONES Y SEGURIDAD SOCIAL
SECRETARÍA DE ESTADO DE MIGRACIONES
SECRETARÍA GENERAL DE INMIGRACIÓN Y EMIGRACIÓN
DIRECCIÓN GENERAL DE INTEGRACIÓN Y ATENCIÓN HUMANITARIA
oberaxe Observatorio Español del Racismo y la Xenofobia

ÍNDICE

Autores:
Nuria Lores y Jesús Migallón. ASOCIACIÓN TRABE

Socios participantes:
Observatorio Español del Racismo y la Xenofobia, CREA-UB, TRABE y Ministerio del Interior

Coordinación del proyecto:
Secretaría General de Inmigración y Emigración - Observatorio Español del Racismo y la Xenofobia (OBERAXE)

Catálogo de publicaciones de la Administración General del Estado
https://cpage.mpr.gob.es

Edita y distribuye: Observatorio Español del Racismo y la Xenofobia
José Abascal, 39, 28003 Madrid
Correo electrónico: **oberaxe@mitramiss.es**
Web: **www.mitramiss.gob.es/oberaxe/index.htm**

NIPO PDF: 854-19-171-6
Diseño y maquetación: Carmen de Hijes

Esta publicación ha sido producida con el apoyo fina pea. Los contenidos de esta publicación son responsabilidad de los socios del proyecto ALRECO y

ÍNDICE

ÍNDICE

INFORME

INFORME 1

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Introducción

ÍNDICE

1. Introducción

El proyecto AL-RE-CO pretende mejorar las capacidades de las autoridades del Estado para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, a fin de diseñar estrategias compartidas frente al discurso motivado por racismo, xenofobia, islamofobia o antisemitismo.

El presente informe se enmarca dentro del **proyecto AL-RE-CO (Discurso de odio, racismo y xenofobia: mecanismos de alerta y respuesta coordinada),** financiado por la Dirección General de Justicia de la Comisión Europea y liderado por el Observatorio Español contra el Racismo y la Xenofobia (OBERAXE) (Secretaría de Estado de Migraciones-Dirección General de Integración y Atención Humanitaria-Ministerio de Trabajo, Migraciones y Seguridad Social).

El proyecto AL-RE-CO pretende mejorar las capacidades de las autoridades del Estado para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, a fin de diseñar estrategias compartidas frente al discurso motivado por racismo, xenofobia, islamofobia o antisemitismo.

Ante el desafío creciente del discurso de odio en línea, las instituciones y autoridades nacionales en España no disponen aún de estrategias de actuación coordinadas. Es preciso reconocer el esfuerzo que han realizado las instituciones españolas en los últimos años para adecuar la normativa y la legislación a estos desafíos, pero en el caso específico de la incitación al odio en línea es necesario reforzar la coordinación y el establecimiento de estrategias de actuación conjuntas.

Esta coordinación y diálogo para la acción común no debe ceñirse al ámbito institucional, sino que debe ser una estrategia que incluya a todos los sectores, especialmente a las organizaciones de la sociedad civil activas en este ámbito y a las propias plataformas en línea, con el objeto de establecer unas bases sólidas para el diálogo y la cooperación entre todas las partes involucradas en la prevención de la incitación del odio en línea.

Dentro de las **fases de trabajo del proyecto** ALRECO se encuentra la **identificación y elaboración de indicadores sobre discurso de odio en la red**. El objetivo es desarrollar un protocolo de actuación que contenga un sistema de indicadores, con criterios de búsqueda, sobre discursos que fomenten el racismo, la xenofobia y el odio en la red[1]. El sistema incluirá también indicadores de alerta temprana que permitan evaluar la intensidad, gravedad, distribución, y potencial impacto del discurso de odio, con el fin de establecer recomendaciones de acción para prevenir posibles incidentes discriminatorios o delitos de odio.

1. Se trabajará específicamente sobre los siguientes motivos de odio: racismo, xenofobia, islamofobia y antisemitismo.

ÍNDICE

Para el marco conceptual en la redacción del presente informe (delitos de odio, racismo, discriminación, etc.) se ha tomado como referencia otros trabajos previos del OBERAXE, en concreto, el *Informe de Delimitación conceptual en materia de delitos de odio.*

En el marco de esta acción se contempla la presente actividad: **la identificación de experiencias y buenas prácticas que se hayan desarrollado en la Unión Europea y que permitan avanzar en el desarrollo del protocolo y del sistema de indicadores. Esta identificación de experiencias servirá de base para el debate y tipificación de indicadores de alerta temprana.**

En base a lo debatido **en la reunión del Kick off del presente proyecto**, celebrada el pasado 20 de diciembre de 2018 en Madrid, *"el informe de BBPP europeas consistirá en la revisión de literatura científica sobre el tema, identificar herramientas que ya se están usando (búsquedas automatizadas de discurso de odio on line en redes sociales –incluida la herramienta que ya dispone el Ministerio del Interior–. Se acuerda incluir también experiencias sobre contra-narrativas, lenguajes excluyentes e incluso-res… (este tema se abordará de forma secundaria, ya que el informe debe centrarse en identificar buenas prácticas sobre indicadores que se usan para identificar el discurso de odio sobre Racismo, Xenofobia, Antisemitismo e Islamofobia). Este informe por tanto alimentará el WP 2 pero también el WP5 (Estrategias Compartidas)".*

En este sentido, el presente informe recogerá no solo las **experiencias en herramientas similares**, sino que incluirá otras experiencias más enmarcadas en el ámbito de la **contranarrativa, la sensibilización,** etc., como se explicará en el apartado 4 del presente informe.

Para el **marco conceptual** en la redacción del presente informe (delitos de odio, racismo, discriminación, etc.) se ha tomado como referencia otros **trabajos previos del OBERAXE**, en concreto, el Informe de ***Delimitación conceptual en materia de delitos de odio***[2] (2018).

Por último, el pasado 3 de julio de 2019 se celebró la primera reunión del Grupo Asesor del Proyecto ALRECO, conformado por instituciones públicas, organizaciones de la sociedad civil y expertos. La finalidad de dicho grupo es velar por el correcto funcionamiento del proyecto. En este sentido, a raíz de la primera reunión se revisó el presente informe y se añadieron dos experiencias y una cita bibliográfica más.

El informe cuenta con cinco partes. En la primera de ellas se describirá **la metodología** empleada para recabar información de las diferentes experiencias. En la segunda se explicarán **los criterios** que se han tenido en cuenta para la selección de las buenas prácticas. En la tercera parte se proporcionará un perfil general de las **características comunes** de las experiencias seleccionadas, y a continuación, se abordará de manera específica la **sistematización de cada una de las prácticas** seleccionadas y, finalmente, se proporcionarán una serie de **conclusiones y recomendaciones** a tener en cuenta para las fases sucesivas del proyecto y, sobre todo, para la elaboración de los indicadores sobre el discurso de odio en la red.

2 http://www.mitramiss.gob.es/oberaxe/ficheros/documentos/InformeConceptualDelitosOdio.pdf

INFORME 2

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Descripción de la metodología empleada

ÍNDICE

2. Descripción de la metodología empleada

Se ha empleado un planteamiento ecléctico e integrador a partir del método de la triangulación, entendido no sólo como la aplicación de distintas metodologías, si no también como una técnica que permite obtener resultados novedosos que se sustentan y enriquecen mediante los diferentes métodos empleados.

2.1. Aspectos preliminares

En línea con la finalidad y objetivos de esta investigación, la propuesta metodológica que se ha empleado para la realización del presente informe recoge una combinación de diferentes perspectivas clásicas de investigación social. La primera limitación con la que se ha contado es el escaso tiempo para la realización del informe (del 2 al 31 de enero de 2019), lo que ha supuesto que no se haya podido realizar una búsqueda exhausiva y un contacto con todas y cada una de las experiencias seleccionadas.

El objeto de estudio, anteriormente descrito y delimitado, se caracteriza por su naturaleza multidimensional y compleja, de ahí que requiera para su comprensión de aquella combinación de instrumentos que se consideren necesarios para captar, en la medida de que los recursos lo permitan, las herramientas existentes para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, por motivos racistas, xenófobos, islamófobos y antisemitas.

Por todo ello (complejidad del fenómeno a estudiar y escasez de tiempo), se ha empleado un planteamiento ecléctico e integrador a partir del método de la triangulación, entendido no sólo como la aplicación de distintas metodologías, si no también como una técnica que permite obtener resultados novedosos que se sustentan y enriquecen mediante los diferentes métodos empleados.

Existen distintas modalidades de triangulación metodológica, aunque para este estudio recurriremos básicamente a tres:

Existen distintas modalidades de triangulación metodológica, aunque para este estudio recurriremos básicamente a tres.

- **Triangulación de datos.** Es el planteamiento más habitual en la práctica de la investigación social. Consiste en la utilización de varias y variadas fuentes de información sobre un mismo objeto de conocimiento, con el propósito de contrastar la información recabada. En el estudio que aquí se presenta se ha recurrido, en las diferentes fases, a diferentes fuentes de información, instituciones, y entidades que aportan datos diversos con lo que se aporta una visión lo más completa e integral posible.

- **Triangulación de investigadores/as.** Corresponde con lo que actualmente se denomina equipos interdisciplinares y consiste en la realización de una misma investigación por un mismo equipo de investigadores/as (procedentes de distintas áreas de conocimiento o especialistas en distintas metodologías), que observan el mismo objeto de estudio desde diferentes puntos de vista, en función de la disciplina científica a la que pertenezcan. El equipo de investigación ha estado formado por dos personas con formación en sociología, antropología y economía; lo que ha permitido aportar visiones complementarias del objeto de estudio, con el fin de afinar el análisis y ajustar, en la medida de lo posible, las conclusiones a la finalidad para la que se plantea la investigación.

- **Triangulación metodológica entre métodos.** Consiste en la combinación de métodos de investigación (no similares) en la medición de una misma unidad de análisis. Con ello se pretenden paliar las limitaciones de cada método, contrarrestándolas con las potencialidades de otros métodos.

En definitiva, las ventajas y potencialidades de esta combinación son múltiples, ya que, en primer lugar, el análisis de diferentes fuentes nos permite captar la evolución y la situación actual y tendencia reciente de las herramientas existentes para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, por motivos racistas, xenófobos, islamófobos y antisemitas, a la vez que nos aporta un conocimiento profundo de los diferentes puntos de vista y prácticas desarrolladas en Europa.

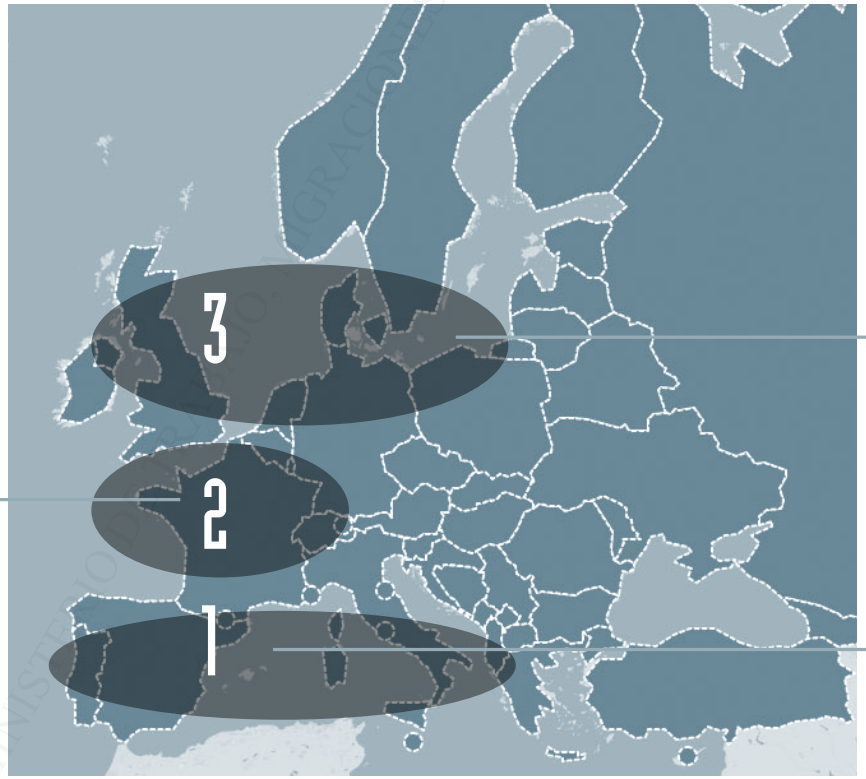
2.2. Técnicas de investigación

Se han desarrollado una serie de herramientas metodológicas a nivel cualitativo y cuantitativo, de cara a obtener unos resultados con mayor espectro de información (nunca con la exhaustividad deseada) y con mayor capacidad explicativa de los datos obtenidos.

En la siguiente tabla se resumen las técnicas de investigación utilizadas:

HERRAMIENTAS	INSTRUMENTOS
ANÁLISIS DOCUMENTAL	Documentos principales: • Proyectos similares • Herramientas ya existentes • Experiencias previas • Artículos académicos y de investigaciones en curso • Análisis de plataformas • Análisis de noticias de prensa relevantes
ENTREVISTAS A PERSONAS CLAVE	• Se ha pedido información a todos los socios del proyecto: OBERAXE, Ministerio del Interior, Universidad de Barcelona (Grupo de investigación CREA), Asociación TRABE. • Se ha contactado con entidades y personas clave de diferentes países europeos (Finlandia, Países Bajos, Austria, Italia, Reino Unido y Grecia). • Se ha contactado con expertos: Asesora de la FRA. • Con entidades sociales: OXFAM.
FICHA - CUESTIONARIO	Desarrollo de una ficha que sistematice la información que es precisa para la valoración de las herramientas.

La diversidad de trayectorias y experiencias migratorias en Europa, así como la pluralidad y complejidad de los distintos modelos de gestión de las migraciones en el continente dificultan la elección de buenas prácticas existentes. Por este motivo, el equipo investigador ha realizado una zonificación de Europa en tres áreas:



Área 2
Europa del Norte
Reino Unido, Países Bajos y Alemania

Área 3
Europa Central
Francia

Área 1
Sur de Europa.
Portugal, España, Italia y Grecia

ÍNDICE

Área 1. Sur de Europa.

Son los países europeos con experiencias migratorias más recientes. Con diferencias significativas entre ellos, especialmente en los últimos años, se pueden observar las tensiones existentes por motivos racistas, xenófobos e islamófobos.

Área 2. Europa Central.

Con una trayectoria de modelos de gestión de la inmigración bien diferenciada de los países del Sur de Europa, con unos vínculos históricos con los países del Magreb, puede haber favorecido el desarrollo de herramientas interesantes a tener en cuenta para el informe de buenas prácticas que se pretende desarrollar en el marco del proyecto ALRECO.

Área 3. Europa del Norte.

En esta zona incluimos países como Reino Unido, Países Bajos o Alemania. No sólo por su situación geográfica, todos estos países se han caracterizado por la aplicación de modelos de gestión de las migraciones multiculturales. Países con una larga trayectoria migratoria y en el desarrollo de políticas de gestión de las migraciones que pueden haber dado lugar al desarrollo de herramientas más consolidadas en el tiempo.

Area 1

- Trayectorias migratorias más recientes
- Políticas interculturales

Area 2

- Trayectorias migratorias amplias
- Políticas de adaptación

Area 3

- Trayectorias migratorias amplias
- Políticas multiculturales

Para lograr cubrir estas tres áreas se ha contactado con diferentes entidades y personas clave, para recabar la máxima información posible en relación con las herramientas existentes para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, por motivos racistas, xenófobos, islamófobos y antisemitas que se estén desarrollando en diferentes partes de la Unión Europea, de cara a ofrecer un resultado completo y diverso que pueda orientar de forma clara las siguientes fases de este proyecto.

ÍNDICE

En la tabla se presenta, de forma resumida, las personas y entidades clave con las que se ha contactado para esta fase:

EXPERTO/ENTIDAD CONTACTADA	PAÍS
European Training and Research Centre for Humans Rigths and Democracy (ETC)	Austria
Centre for European Constitutional Law (CECL)	Grecia
Universidad de Milán	Italia
Ministry of Justice of Finland. Anti-discrimination and Fundamental Rights Team	Finlandia
Bradford Hate Crime Alliance	Reino Unido
Department of European and International Affairs/City of Utrecht	Países Bajos
Rosa Bada, Board Member FRA Advisory	Experta Europea
Jose Camacho-Collados, Universidad de Cardiff	Reino Unido
Juan Carlos Pereira Kohatsu	TFM de la Universidad Carlos III España
SOCIOS DEL PROYECTO	**PAÍS**
Observatorio Español contra el Racismo y la Xenofobia (OBERAXE) (Secretaría de Estado de Migraciones- Ministerio de Trabajo, Migraciones y Seguridad Social)	España
Oficina nacional de lucha contra los delitos de Odio (Secretaría de Estado de Seguridad, Ministerio del Interior)	España
Universidad de Barcelona (Grupo de investigación CREA)	España
Asociación TRABE	España

Finalmente se ha elaborado una ficha – cuestionario para recoger información de cara al posterior desarrollo de indicadores contemplado en el proyecto. La ficha-cuestionario ha permitido, con carácter sistemático y homogéneo, la comparación entre experiencias formativas y su sistematización, tal y como se recoge en el apartado 5 del presente informe.

INFORME 3

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Criterios para la selección de las buenas prácticas y experiencias

ÍNDICE

3. Criterios para la selección de las buenas prácticas y experiencias

Fue necesario partir de una definición consensuada por los socios del proyecto de lo que se entiende por buena práctica en relación con las herramientas y experiencias en el ámbito del discurso de odio en línea.

En este sentido, y como resultado de la revisión bibliográfica realizada, en el marco del proyecto ALRECO se propone como definición de buena práctica la siguiente:

> Aquella actuación, metodología o herramienta desarrolladas en Europa, en el ámbito del discurso de odio en línea, que ha mostrado su capacidad para introducir transformaciones con resultados positivos en la identificación, análisis, monitorización y/o evaluación del discurso de odio en línea por motivos racistas, xenófobos, islamófobos y antisemitas.
>
> En concreto, se valorarán aspectos como la trayectoria de la experiencia, que se haya implementado y que haya obtenido algún tipo de resultados en la práctica, el disponer de mecanismos de evaluación, la heterogeneidad del conjunto de experiencias en cuanto al agente promotor y a los beneficiarios de la experiencia (institucional, universidad, impulsadas por el tercer sector, etc.), que cuenten con mecanismos de coordinación.

Dado que este informe es un primer producto del WP 2 del Proyecto ALRECO, se ha optado por **incluir aquellas herramientas, actuaciones, metodologías que pudieran servir para las fases sucesivas** contempladas en el WP2:

WP 2 PROYECTO ALRECO:

PROTOCOLO Y SISTEMA DE INDICADORES CONTRA EL RACISMO, LA XENOFOBIA Y EL DISCURSO DE ODIO

ELABORACIÓN INDICADORES (FASE 1):	Tipificación de factores clave para el análisis del discurso de odio en internet y en las redes sociales: identificación de palabras clave, frases, conceptos, redes sociales (especialmente Facebook y Twitter). Que se identifiquen como discurso de odio en al menos cuatro ámbitos: racismo, xenofobia, islamofobia y antisemitismo, así como páginas web o perfiles determinados de redes.
PRIMER TALLER-TRABAJO CON TODOS LOS SOCIOS DEL PROYECTO.	Revisión de la metodología propuesta y del documento sobre los criterios de búsqueda y elaboración de propuestas y recomendaciones al mismo.
ELABORACIÓN DE INDICADORES (FASE 2):	Construcción de un sistema de indicadores que permitan evaluar la intensidad, gravedad, distribución, y potencial impacto del discurso de odio, con el fin de establecer recomendaciones de acción en función de los mismos y prevenir posibles incidentes discriminatorios o delitos de odio. Este sistema de indicadores es el que servirá de base para el desarrollo de la herramienta

INFORME 4

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Características generales de las experiencias seleccionadas

ÍNDICE

4. Características generales de las buenas prácticas y experiencias seleccionadas

Con la metodología empleada explicada en el apartado 2 del presente informe, se ha obtenido información de un total de 53 experiencias en el periodo comprendido entre el 9 y el 30 de enero de 2019. El grupo asesor del proyecto aportó información adicional sobre 2 experiencias más, por lo que se cuenta con un total de 55 experiencias.

De las 55 experiencias se han seleccionado 20, tal y como se recoge en el apartado siguiente del presente informe. Esta selección se ha basado en los criterios consensuados y expuestos en el apartado anterior.

En cuanto a los ámbitos de las experiencias seleccionadas cabe destacar que la mayoría de ellas combinan diferentes aspectos, es decir, no hay apenas herramientas "puras" si no que, más bien, tienen una parte de herramienta y otra parte de contranarrativa, algunas de sensibilización, de formación. De las 20 experiencias seleccionadas y apenas hay dos que se pueden considerar herramienta "pura".

En cuanto a las zonas geográficas a las que se refieren las experiencias seleccionadas, si bien el ámbito del informe es europeo, se ha incluido una herramienta de Estados Unidos, porque era especialmente relevante para el objetivo del proyecto ALRECO y porque es difícil delimitar las zonas geográficas de impacto. Por ejemplo, otra de las herramientas seleccionadas se ciñe al espacio "hispanohablante", que trasciende el espacio europeo. El resto de las experiencias abordan diferentes países, al ser experiencias promovidas en consorcio por diferentes socios. Una cuestión a tener en cuenta no es sólo el país donde se promueve la experiencia sino el idioma en el que es efectiva la experiencia (p.ej no disponemos de ninguna experiencia árabe pero si que trabajen en ese idioma). En definitiva, se han recogido experiencias de los siguientes países: España, Francia, Italia, Reino Unido, Grecia, Alemania y Estados Unidos.

En cuanto a los colectivos a los que se refieren las diferentes experiencias seleccionadas, la mayoría de ellas están dirigidas a minorías migrantes o étnicas y diferentes opciones religiosas (musulmanes y judíos fundamentalmente). Algunas combinan diferentes criterios y se dirigen a colectivos vulnerables (LGTBI, pueblo gitano, etc.).

En cuanto al tipo de promotores de las diferentes experiencias seleccionadas, básicamente son tres: Universidades/ámbito Académico, instituciones públicas y ONG. En cuanto a la financiación son en su mayoría financiadas mediante fondos públicos, si bien se han incluido algunas experiencias financiadas directamente por empresas como Google o Facebook.

En el siguiente apartado se han sistematizado las 18 experiencias seleccionadas. Si bien, no se ha podido profundizar en cada una de ellas, no obstante, servirá de punto de partida para la elaboración de las siguientes tareas contempladas en el WP2: protocolo y sistema de indicadores contra el racismo, la xenofobia y el discurso de odio.

INFORME

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Sistematización de las experiencias seleccionadas

ÍNDICE

5. Sistematización de las experiencias seleccionadas

Nombre	**Rastreador de odio. (Monitoreo y clasificador en español de discurso de odio en la red social Twitter.)**
Dirección web-contacto	Uso interno de la Oficina de Delitos de Odio - Secretaría de Estado de Seguridad de España, Ministerio del Interior. La herramienta trabaja a nivel local en aquellos PC´s donde es instalada, los servidores que acceden a los contenidos de la red social Twitter y proveen la información se encuentran actualmente en la Universidad Autónoma de Madrid.
Quién la promueve	La herramienta es fruto de un TFM de la Universidad Carlos III de Madrid cuyo autor es Juan Carlos Pereira Kohatsu, bajo la tutoría de Lara Quijano Sánchez, Álvaro Ortigosa y Miguel Camacho-Collados.
Quién la financia	No está financiada. Es una investigación en el marco universitario de un TFM. Actualmente están en periodo de valoración los recursos que se destinan a esta iniciativa.
Breve descripción [1]	Dada la creciente necesidad de monitorear los mensajes de odio en las redes, el objetivo principal de este TFM es modernizar y mejorar el rendimiento de los clasificadores actuales de odio y desarrollar una herramienta de análisis de datos que analice el estado de odio en Twitter en España e identifique los tweets dañinos, las tendencias de odio y la construcción de otros sentimientos negativos. Objetivo general: Rastrear en un número determinado de mensajes de la red social Twitter el contenido que pudiera considerarse discurso de odio. Objetivos específicos: Analizar los mensajes y símbolos que aparecen en los tweets y que podrían contener odio, analizar la interacción entre usuarios a través de estos mensajes en twitter. Analizar las comunidades digitales relacionando la gente que interactúa, y el impacto social del mensaje que contiene. Realizar un análisis grafico de toda esta información mediante una interfaz intuitiva.

Nombre	**Rastreador de odio. (Monitoreo y clasificador en español de discurso de odio en la red social Twitter.)**
Breve descripción [2]	Descripción de la metodología empleada: Data Science, la herramienta, fue desarrollada en base al procesamiento del Lenguaje natural, utiliza algoritmos que forman una red neuronal, miden y clasifican los contenidos. Se estima que al día se escriben unos 100 000 mensajes tóxicos y que pasan los filtros que la propia red social impone a sus usuarios. Esta herramienta detecta el 24% de mensajes con odio de una muestra de 6.000 tweets. Descripción de los contenidos: el programa tiene una interfaz sencilla para el usuario, se visualizan en pantalla los mensajes ordenados en series de 10 con los porcentajes de contenido de odio. Estos pueden ser seleccionados para su análisis independiente o copiar su enlace para ser vistos directamente en la red social. Posteriormente tenemos los apartados gráficos que permiten visualizar las comunidades de los propios usuarios de redes mediante el grafismo de nodos unidos por flechas. Este gráfico es interactivo y de fácil comprensión para analizar los mensajes enviados y recibidos. Principales discursos de odio identificados por la herramienta: discurso emitido por personajes de carácter público de diferentes ámbitos como el político, social, deportivo o de entretenimiento que reciben gran cantidad de estos mensajes. A su vez, los productores de los mensajes suelen ser los mismos. Los contenidos suelen ser ofensivos para quien los recibe o bien para la comunidad o colectivo que los representa. Hasta donde se ha podido probar y realizar muestreos, se ha detectado contenido tóxico en cuentas privadas. Factores que se han identificado como posibles promotores del auge de los discursos de odio (noticias, bots, …): se ha observado que está directamente relacionada una noticia y un hecho social con el aumento puntual de discurso de odio. Se ha observado también que los bots generan un discurso de odio que es repetido por los mismos usuarios de la red social. Además, hay un número determinado de pequeñas comunidades que generan o reciben gran cantidad de mensajes con contenido tóxico o de odio.

Nombre	**Rastreador de odio. (Monitoreo y clasificador en español de discurso de odio en la red social Twitter.)**
Trayectoria	El TFM se finalizó en septiembre de 2018 y en la actualidad está siendo empleado en modo interno por la Oficina de Delitos de Odio.
A quién se dirige	Es una herramienta destinada para su uso inicial por la Oficina de Delitos de Odio - Secretaría de Estado de Seguridad de España, Ministerio del Interior. La herramienta reconoce 6 tipos de delitos de odio: por motivos étnicos, por género, discapacidad, ideología política y religión.
Puntos positivos	La herramienta tiene varias aplicaciones en relación con el estudio de la sociedad y sus tendencias, orientando al usuario en la prevención del fenómeno del discurso de odio mediante la aplicación de políticas y contranarrativa. Además, puede ser utilizado por la policía ayudando a las investigaciones de ciberdelincuencia. Los trabajos realizados hasta la fecha dan pie para mejorar y generar nuevos clasificadores y centralizar las BBDD a otras instituciones que pudieran utilizarla y mejorar también los patrones de búsqueda.
Carencias	Es un sistema que esta alimentado en idioma español obviando el resto de las lenguas que conviven en España. La herramienta no puede diferenciar los Tweets que provengan de otros países de habla hispana ya que el proveedor de la plataforma no identifica la ubicación geográfica ni horaria.
Resultados obtenidos	Exactitud en su clasificación del 80,15%. Probabilidad del 77,57% en que el clasificador acierte al decir que un mensaje contiene odio y un 33,75% de tweets clasificados finalmente como odio.

Nombre	**Rastreador de odio. (Monitoreo y clasificador en español de discurso de odio en la red social Twitter.)**
Principales conclusiones	La herramienta permite analizar la evolución de la tipología penal del discurso de odio, su durabilidad y los patrones que se usan para realizar esta conducta, los discursos de odio deben ser estudiados y observados, ya que actualmente las redes sociales son prioritarias para la vida social por la carga de información que contienen. En estos momentos este "desarrollador" filtra en "contienen odio" y "no contienen odio" mediante un sistema de *Machine Learning* y que no sustituye la mano del hombre como algo determinante en su última instancia. Permite conocer las comunidades "Twitteras" que generan un discurso de odio y nos hace anticiparnos a este, conocer el perfil humano y social de los usuarios que generan discurso de odio y pone de manifiesto la capacidad que tiene esta herramienta para predecir un discurso de odio que puede ser provocado o con ocasión de un evento o hecho social.
Género	La variable de género es una de las seis analizadas. La herramienta detecta todos los insultos como *feminazi*. Se está trabajando en nuevas iniciativas complementarias de detección y clasificación de odio misógino.
Observaciones generales	Es una herramienta muy reciente que aún está en fases de prueba y testeo para incorporar mejoras.

Nombre	**Somos Más**
Dirección web - contacto	www.somos-mas.es
Quién la promueve	La campaña Somos Más se realiza con la colaboración de YouTube, el Gobierno de España (específicamente a través del Ministerio de Justicia); el Ministerio del Interior (Secretaría de Estado de Seguridad y CITCO); el Ministerio de Educación, Cultura y Deporte (Centro Nacional de Innovación e Investigación Educativa); el Ministerio de Empleo y Seguridad Social (Secretaría General de Inmigración y Emigración, OBERAXE); y el Ministerio de Sanidad, Servicios Sociales e Igualdad (Secretaría de Estado de Servicios Sociales e Igualdad, a través del Instituto de la Mujer y para la Igualdad de Oportunidades y de INJUVE); la Red Aware (Alliance of Women Against Radicalization and Extremism); FeSP-UGT (Aula Intercultural) y la ONG Jóvenes y Desarrollo.
Quién la financia	Google.org a través de la iniciativa global YouTube Creators for Change.
Breve descripción	La campaña "Somos Más, contra el odio y el radicalismo" tiene como objetivo: prevenir y sensibilizar sobre el discurso del odio y la radicalización violenta. Somos Más se divide en 2 ejes de trabajo principales: • Formación: talleres dirigidos a chicos y chicas en escuelas y centros de todo el país que incluyen el desarrollo de material didáctico y juegos de rol para facilitar la comprensión por parte de alumnos, padres y profesores. • Sensibilización: campaña de comunicación que incluye, entre otras cosas, colaboraciones con creadores que ayudarán a amplificar mensajes positivos a través de sus vídeos.

Nombre	**Somos Más**
Trayectoria	Campaña iniciada en 2018. En enero de 2019 se han presentado los resultados de la primera edición entre los que destacan la participación de 235 centros educativos y la formación de 28.238 jóvenes entre 14 y 18 años en el primer eje de trabajo y más de 43M de acciones de sensibilización en redes sociales en las que la colaboración de los creadores youtubers tienen un papel fundamental. En cada edición de la campaña son invitados a participar 8 creadores de perfil diverso y comprometidos/as en la lucha contra el discurso de odio en internet.
A quién se dirige	La campaña está dirigida principalmente a jóvenes de entre 14 y 20 años de edad de todo el país. Asimismo, en la página web del proyecto están disponibles los materiales didácticos con el fin de alcanzar también a padres, tutores, escuelas y otros actores importantes en el ámbito educativo.
Observaciones generales	Se va a analizar más esta experiencia en una segunda fase de trabajo.

Nombre	**CiberHache**
Dirección web contacto	http://ciberhache.com
Quién la promueve	Centro CRÍMINA para el estudio y prevención de la delincuencia de la Universidad Miguel Hernández.
Quién la financia	Referencia: DER2014-53449-R Ministerio de Economía y Competitividad.
Breve descripción	El proyecto CiberHache, llevado a cabo por el Centro CRÍMINA para el estudio y prevención de la delincuencia (UMH), estudia la incitación a la violencia y discurso del odio en Internet, considerando el alcance real del fenómeno, tipologías, factores ambientales y límites de la intervención jurídica. El objetivo general consiste en obtener una imagen lo más cercana a la realidad de los fenómenos de la incitación a la violencia y el discurso del odio en España. Una imagen empíricamente contrastada, que se aleje tanto de exageraciones mediáticas como de banalizaciones sobre la gravedad de alguna de las conductas. Ello obliga a una esencial categorización de las, muy diferentes entre sí, formas de violencia en Internet, según la intensidad de la incitación, su ámbito, los distintos intereses puestos en riesgo, etc. El segundo gran objetivo supone establecer un marco jurídico de respuesta a las diferentes formas de incitación a la violencia y de comunicación de odio en Internet proporcional a la distinta gravedad de cada una de las tipologías, a la vez que respetuoso con los derechos y principios fundamentales de un Estado Democrático de Derecho como el que plasma nuestra Constitución.

Nombre	**CiberHache**
Concepto de hate speech en el marco de la investigación	A los efectos de este trabajo la comunicación violenta engloba cualquier forma de expresión que pueda conceptuarse como violenta, independientemente de que se lleve a cabo por motivos discriminatorios. La violencia constituye, pues, el elemento definitorio de este tipo de comunicación, integrando en la misma tanto la violencia física anunciada, incitada, deseada, justificada o valorada positivamente, como la violencia moral, aquella que resulta de la causación de un daño no físico o de una ofensa a intereses morales dignos de tutela de personas concretas o de una colectividad. El hate speech o discurso de odio, pues, tal como ha sido definido, podría considerarse comunicación violenta, en cuanto que la incitación a la violencia lo es en sentido de violencia física, y la incitación al odio o a la discriminación constituye una forma de violencia moral. Pero hay otras formas de comunicación violenta distintas al hate speech, concretamente: a) todas las formas de incitación (directa o indirecta), o amenaza específica de causación de violencia física que no sean por razón discriminatoria o de grupo; b) toda ofensa o daño al honor o a la dignidad de personas concretas y c) todo comportamiento que pueda considerarse ofensivo o vejatorio para la sociedad aunque no vaya dirigido a una persona en concreto.

Nombre	**CiberHache**
Metodología	Se parte de la metodología que se aplicó a raíz del atentado al semanario francés "Charlie Hebdo"(255.674 tuits a lo largo de seis días). Se establecen cinco criterios: 1. Insultos graves 2. Referencia en positivo a la violencia 3. Atribución a personas en concreto de expresiones injuriosas, humillación pública o la imputación de hechos delictivos o ilícitos graves 4. Desprecio o expresión de odio hacia grupos determinados 5. Aquellas expresiones especialmente desagradables y de muy mal gusto sobre sucesos que causen grave dolor. A través de estos criterios, los evaluadores, previamente entrenados, cribaron y seleccionaron los mensajes que cumplían con alguna de las formas que se expresa en cada catalogación (ver cuadro más abajo). El triaje les condujo a que 2.274 tuits eran violentos (46,9%) y de odio (43,8%) o ambos a la vez (9,3%). En términos generales, de todas las conversaciones que se generaron a través de la red social señalada en torno a tres hashtags seleccionados, solo aproximadamente el 2% era comunicación violenta o de odio, el resto fue comunicación neutral.
Observaciones generales	Numerosos materiales por internet de interés. La monografía final del proyecto (no disponible online): Cometer delitos en 140 caracteres: el Derecho Penal ante el odio y la radicalización en Internet. Fernando Miró Llinares (dir.). El informe nacional de CIBERHACHE está bajo petición en el mail **secretaria@crimina.es**
Links de interés	**https://youtu.be/KKiV1D90r0o?list=PL9smxsRfJ135PgFArkmph7NI-hI_XvFBa**

Nombre	**cibeRespect**
Dirección web - contacto	http://www.ecosdosur.org/ciberespect Responsable de comunicación: Natalia Monge.
Quién la promueve	Ecos do Sur en partenariado con Sos Racisme, Institut de Drets Humans de Catalunya y United Explanations.
Quién la financia	Obra Social La Caixa.
Breve descripción	La capacitación y la lucha contra el discurso del odio xenófobo es el núcleo de acción del proyecto cibeRespect. Propone combatir el impacto que el discurso del odio en la Red puede ejercer sobre la opinión pública a través de un nuevo modelo de intervención que articula el espacio físico y el virtual. Para esto, proyecta la acción local y las relaciones de cercanía hacia el entorno global digital, inspirando y acompañando el nacimiento de un nuevo agente social: el/la ciberactivista llamado/a a liderar la lucha contra el discurso del odio en Internet. CibeRespect incluye las siguientes acciones: • Seguimiento y análisis de discurso de odio xenófobo en Internet. • Formación de mesas de trabajo. • Elaboración de materiales de contraargumentación: manuales, infografías, página web. • Curso de formación a ciberactivistas. • Red de ciberactivistas. • Intervención en foros y RRSS de Medios de Comunicación.
Trayectoria	Desde 2017. No se tiene constancia de que continúe su actividad después de la financiación de La Caixa (el importe fue de 60.000 euros).
Observaciones generales	Actualmente esta experiencia no está vigente. El proyecto sigue en funcionamiento de una manera manual, no disponen de herramientas informáticas sofisticadas.

Nombre	**Observatorio PROXI**
Dirección web - contacto	http://www.observatorioproxi.org Anna Palacios, Institut Català de Drets Humans.
Quién la promueve	Institut de Drets Humans de Catalunya, United Explanations, Plataforma de ONG de Acción Social.
Quién la financia	EEA Grants.
Breve descripción	El Proyecto Online contra la Xenofobia y la Intolerancia en Medios Digitales es una iniciativa de diversas entidades de derechos humanos para luchar contra el discurso de odio en Internet. El también llamado ciberodio es un fenómeno creciente en toda Europa contra el que urge una actuación global. Concretamente, en España, la crudeza de las medidas de recorte del gasto público y la presencia de persistentes actitudes xenófobas en la población son el caldo de cultivo ideal para que se propague el discurso de odio y para que ideas xenófobas e intolerantes se difundan. Este contexto, sumado a la utilización creciente del espacio online y de sus nuevos canales de difusión e intercambio, favorece la propagación del discurso de odio a la vez que dificulta la identificación de sus responsables y su contención. El Observatorio PROXI tiene un doble objetivo. Por una parte, monitorear y analizar el discurso de odio en los comentarios de los medios digitales y noticias seleccionadas. Por otra, realizar intervenciones directas para contrarrestar los argumentos que utiliza el ciberodio y para hacer presente un discurso alternativo. Para ello, se tiene una estrategia de intervención en tres líneas paralelas: a) Identificar y analizar el discurso de odio en los hilos de comentarios de noticias sobre inmigrantes y población gitana. b) Contraargumentar el discurso de odio con la elaboración de un discurso alternativo basado en los derechos humanos. c) Prevenir el discurso de odio en internet a través de la formación de jóvenes internautas.

Nombre	**Observatorio PROXI**
Trayectoria	Proyecto que ha sido financiado por los EEA Grants y que ha desarrollado su actividad hasta 2015. No continúa realizando el monitoreo. Su último informe accesible es del 2015 aunque aún es posible descargarse el software *Proxi Comment Analyzer* que es un software desarrollado específicamente para la realización de este proyecto, tiene Licencia GPL3, conocido como software libre (Free Software). Por tanto, se concede libertad absoluta para su utilización, modificación, lectura y uso.
A quién se dirige	Grupos objeto de estudio: población migrante y población gitana. En la situación de crisis económica, tanto migrantes como población gitana son percibidos como potenciales competidores por los recursos sociales y laborales. Por ello, estos dos colectivos se consideran vulnerables y especialmente expuestos a ser objeto de ataques y discriminación. Medios digitales analizados. El monitoreo de los hilos de comentarios se lleva a cabo mediante el rastreo de noticias en los tres diarios generalistas online de mayor audiencia. Estos son: El País, El Mundo y 20 Minutos. Estos medios digitales ofrecen un campo de estudio de la gestación y propagación del discurso de odio mediante la recolección de los comentarios que genera. El análisis del contenido de los comentarios permite, además, identificar los argumentos utilizados en el discurso de odio, paso previo necesario para lograr desmontarlos. Selección de las noticias. Se seleccionan noticias que tratan del colectivo migrante o del colectivo gitano en los tres medios digitales. La selección se realiza diariamente utilizando la herramienta del sistema de alertas de Google (Google Alerts), y se complementa con un seguimiento cualitativo para localizar noticias que no hayan sido detectadas por el sistema de alertas. Monitoreo de comentarios. Para facilitar la tarea de monitoreo y automatizar la recogida de datos, este proyecto ha desarrollado un software denominado Proxi Comment Analyzer que permite el seguimiento y volcado de los comentarios identificados en los diarios digitales objeto de estudio. (Este software se puede descargar desde la web del proyecto, yendo al apartado "Actúa"). Categorización de los comentarios. El Observatorio PROXI contempla cinco categorías para los distintos grados de "discurso de odio", dentro de la definición amplia de "discurso de odio" utilizada en este proyecto: odio, estereotipo y prejuicio, rumor, argumento trampa y argumento antiinmigración/antigitano de baja intensidad.

Nombre	**Observatorio PROXI**
Metodología Categorización de los comentarios [1]	1.Odio Comentarios que contienen lenguaje insultante y degradante y/o incitan a la violencia. Ejemplo: "De la misma manera que entran ilegalmente se les debe expulsar inmediatamente. Ahora tocará quitar el pan de la boca a los españoles para dar de comer a esta escoria sarnosa y ellos te lo agradecerán imponiendo su cultura musulmana en tu casa". [Comentario a la noticia "Interceptan dos pateras, con 75 inmigrantes, en aguas españolas en las últimas doce horas", publicada en 20 Minutos el 16 de junio de 2014]. 2. Estereotipos y prejuicios Comentarios que comunican ideas muy simplificadas sobre miembros de una comunidad, una generalización que no tiene en cuenta diferencias individuales. Cuando estos estereotipos son negativos, también hablamos de prejuicios. Ejemplo: "no es discriminación, sino, ante todo, un problema de higiene. ¿Dónde hacen sus necesidades vitales? ¿Dónde están los maridos de esas señoras con niños? ¿Robando en otro sector de París o en el metro? ¿O simplemente llegan a la hora de sacarle el dinero recaudado durante el día?". [Comentario a la noticia "Una comisaría de París recibe la orden de expulsar sistemáticamente a los gitanos" publicada en El País el 15 de abril de 2014]. 3. Rumor Comentarios que constituyen declaraciones basadas en falsedades sobre personas o grupos que se difunden de una persona a otra sin que se demuestre su veracidad. Ejemplo: "Otros que van a tener más derechos y ayudas que los españoles". [Comentario a la noticia "Interceptan dos pateras, con 75 inmigrantes, en aguas españolas en las últimas doce horas" publicada en 20 Minutos el 16 de junio de 2014]. 4. Argumento trampa Comentarios que no tienen respuesta posible porque niegan cualquier posibilidad de debate. Sitúa el debate en un escenario que no es realista. Ejemplo: "¿Cuántos tienes en tu casa acogidos?". [Comentario a la noticia "España y Marruecos evitan la entrada a Melilla de unos mil subsaharianos" publicada en El País el 14 de junio de 2014"].

Nombre	**Observatorio PROXI**
Metodología Categorización de los comentarios (2)	5. Discurso antiinmigración o antigitano de baja intensidad Comentarios que no encajan adecuadamente en las categorías de discurso del odio (odio, estereotipo, rumor y argumento trampa), pero tampoco se pueden considerar como neutros, porque, o bien dejan entrever un discurso crítico y negativo con la inmigración o la población gitana, o bien contienen un discurso de crítica indirecta, es decir son críticos con las instituciones o entidades que protegen los derechos de los inmigrantes y/o los gitanos, tales como las ONGs, jueces, u otras instituciones estatales o europeas. Ejemplo: "Menudos jueces y menudas ONG así nos luce el pelo a los españoles. De aquí a poco nos ponen la frontera en los Pirineos otra vez". [Comentario a la noticia "Un juez indaga maltratos de policías marroquíes a inmigrantes en Melilla" publicada en El País el 7 de agosto de 2014]. 6. Discurso alternativo Comentarios que hacen presente en los foros de los medios digitales un discurso alternativo al del discurso de odio. Se trata de comentarios que contienen una base de respeto a los Derechos Humanos, o fundados en el derecho de la gente a progresar, a una vida mejor, así como comentarios que matizan afirmaciones inexactas, desmienten rumores, o refutan datos incorrectos de otros comentarios, Ejemplo: "Te hablo de racismo porque comparas y sigues comparando a todo un pueblo con unos individuos determinados (también hay españoles que hacen eso que dices) y los comparas con adjetivos como corruptos solo por ser de una raza". [Comentario a la noticia "Una comisaría de París recibe la orden de expulsar sistemáticamente a los gitanos" publicada en El País el 15 de abril de 2014]. 7. Comentario neutro Comentarios que no pueden clasificarse en ninguna de las otras categorías. La inclusión de esta categoría se hace con el objetivo de poder analizar el porcentaje de comentarios del resto de categorías sobre el total a la hora de procesar los resultados de la categorización de comentarios. Ejemplo: "ya está la censura del 'diario de izquierdas`… y lo peor es que me borra los mensajes un becario explotado que le pagarán cuatro duros….". [Comentario a la noticia "Una comisaría de París recibe la orden de expulsar sistemáticamente a los gitanos" publicada en El País el 15 de abril de 2014].

Nombre	**Observatorio PROXI**
Observaciones generales	La iniciativa dispone de un software libre para descargar (ver link más abajo). El proyecto ya no sigue pero están dispuestos a contar su experiencia para el desarrollo de la herramienta que ya tenían y, sobre todo, de cómo realizaron la categorización de las palabras clave. Las categorías están en la ficha pero puede ser interesante para ALRECO hablar con ellos del proceso. Pueden facilitar el contacto con United Explanations que estaban en el consorcio del proyecto y fueron los responsables de esta categorización. Nos parece interesante para esta segunda fase del proyecto.
Links de interés	http://www.observatorioproxi.org/images/pdfs/INFORME-proxi-2015.pdf http://www.observatorioproxi.org/index.php/observa/metodologia

Nombre	**Be the Key**
Dirección web - contacto	https://www.facebook.com/pg/bethekeybarcelona/about/?ref=page_internal
Quién la promueve	Un grupo de alumnos de la Facultad de Comunicación y Relaciones Internacionales Blanquerna (Barcelona).
Quién la financia	Apoyo y financiación de Facebook.
Breve descripción	Be the Key (#BeTheKey) es una campaña que contrarresta el extremismo, el discurso del odio y la islamofobia. Con sede en el Raval, Barcelona.
Trayectoria	Tras los atentados del 17 de agosto de 2017, que golpearon el barrio del Raval, donde está ubicado su centro de estudios, un grupo reaccionó para combatir las actitudes intolerantes y trasladar mensajes positivos a través de una campaña en Facebook. Desde entonces, lo hacen en plataformas como Facebook, Instagram y Twitter, y con mensajes publicados en catalán, castellano, inglés, árabe, francés, italiano y griego. Ya han recibido contenidos de Estados Unidos, Canadá, Chile, Alemania, Italia, Bélgica, Eslovenia, Marruecos y España. A la campaña se han unido instituciones como el Ayuntamiento de Barcelona, el Instituto Europeo del Mediterráneo (IEMed) y la Fundación Anna Lindh.
Observaciones generales	Se trata de una campaña en Facebook que cuenta con 1.027 seguidores.

Nombre	**http://oficialrewind.com**
Dirección web - contacto	http://www.oficialrewind.com
Quién la promueve	Sin datos en la web.
Quién la financia	Sin datos en la web.
Breve descripción	Rewind tiene como objetivo inspirar un uso más consciente de las redes sociales, así como fomentar la respuesta a los mensajes que generan odio. El objetivo es conseguir un internet con opiniones basadas en argumentos. Alentar a las personas a que piensen antes de escribir y vuelvan atrás antes de darle a la tecla enviar.
Trayectoria	No es posible determinar desde cuándo están en funcionamiento.
A quién se dirige	Comunidad hispanohablante en redes sociales.
Observaciones generales	Disponen de una cuenta en twitter con 467 seguidores pero su último mensaje es de mayo de 2017. En Facebook tienen cerca de 12.000 seguidores pero su último mensaje es de junio de 2017.

Nombre	**Monitoring and Detecting OnLine Hate Speech (MANDOLA)**
Dirección web - contacto	www.mandola-project.eu
Quién la promueve	Foundation for Research and Technology - Hellas (FORTH), Aconite Internet Solutions, International Cyber Investigation Training Academy, Inthemis, Universidad Autónoma de Madrid, Universidad de Chipre, Universidad de Montpellier.
Quién la financia	Este proyecto está financiado por el Rights, Equality and Citizenship (REC) Programme de la Comisión Europea.
Breve descripción	El proyecto MANDOLA pretende mejorar la comprensión de la prevalencia y difusión del discurso de odio en línea y capacitar a los ciudadanos para que monitoreen y denuncien el discurso de odio. Los objetivos de este proyecto son: • Monitorear la difusión y la penetración del discurso en línea relacionado con el odio en Europa y en los Estados miembros utilizando enfoques de big data. Al mismo tiempo que se investiga la posibilidad de distinguir, entre los contenidos monitoreados, entre discursos potencialmente ilegales relacionados con el odio y el odio potencialmente no ilegal. • Proporcionar a los responsables de la formulación de políticas de información procesable que pueda utilizarse para promover políticas que mitiguen la propagación del discurso de odio en línea. • Proporcionar a los ciudadanos herramientas útiles que puedan ayudarlos a trabajar contra el discurso de odio. • Transferir las mejores prácticas entre los Estados miembros. • Establecer una infraestructura de informes que conectará a los ciudadanos preocupados con las fuerzas policiales y que permitirá informar sobre el discurso ilegal relacionado con el odio.
Trayectoria	Proyecto en funcionamiento desde 2017.
Observaciones generales	Dispone de una web asociada para reportar casos de discurso de odio online pero los recursos son limitados. Indican a quién se puede denunciar un caso de discurso de odio en cada país de la UE y tienen un listado de entidades para cada país.

Nombre	**Words Are Stones**
Dirección web - contacto	https://www.wordsarestones.eu/
Quién la promueve	Istituto Europeo per lo sviluppo economico (ISES) – Italia Como proyecto europeo, tiene distintos socios que aparecen en la web.
Quién la financia	Esta web ha sido financiada por el programa REC de la Unión Europea.
Breve descripción [1]	WORDS ARE STONES pretende intervenir para procesar casos de discursos de odio en línea y cooperar con las empresas de IT y los medios de comunicación, para combatir el discurso de odio manifiestamente ilegal y promover contra-narrativas que emanan de la sociedad civil mediante: • Organización de actividades capaces de apoyar a la sociedad civil en el desarrollo de contra-narrativas en línea. • Organización de actividades capaces de apoyar la alfabetización mediática a través de la capacitación y la difusión de datos. • Sensibilización de los medios para promover la diversidad y la tolerancia. El objetivo general del proyecto es combatir el racismo y la discriminación en su expresión en línea del discurso de odio equipando a jóvenes estrategas / administradores de redes sociales, bloggers, activistas en línea, youtubers ... y a los jóvenes en general, con las competencias necesarias para reconocer y actuar en contra de tales violaciones de derechos humanos. En el centro de la filosofía del proyecto está la idea de que el espacio en línea es un espacio público, los derechos humanos se deben aplicar igual que en el resto de la sociedad. Al ser el proyecto implementado en Italia, Bulgaria, Grecia, Rumania, España, Hungría, República Checa, Lituania, tendrá un impacto en los países que experimentan de primera mano el discurso de odio. Por lo tanto, se contribuirá a la creación de una red de la UE para fomentar la cooperación internacional en el campo y promover un intercambio continuo de mejores prácticas.

Nombre	**Words Are Stones**
Breve descripción (2)	En particular, las actividades del proyecto son: • La organización de un curso de formación en Italia para jóvenes estrategas / gestores de redes sociales, bloggers, actividades en línea, youtubers y su replicación en cada país participante. • La organización de una "campaña de medios juveniles" con actividades locales dentro y fuera de línea. Los materiales producidos servirán de base para el "Premio de WORDS ARE STONES", un momento de celebración en Europa mediante el cual es posible informar y votar los mejores casos de gestión del discurso de odio, y la mejor conducta de los usuarios de Internet para un internet más inclusiva.
A quién se dirige	Los grupos a los que se dirige del proyecto son: • Jóvenes bloggers, activistas de redes sociales, gestores comunitarios, moderadores, de 18 a 30 años de edad, con capacidad probada para movilizar a los jóvenes en línea. • Jóvenes de 14 a 25 años. Los hemos identificado como grupo objetivo porque los jóvenes ahora son de la generación "Web 2.0": los que usan Internet están más familiarizados con sus diferentes aspectos y son completamente capaces de hacer uso de ellos. Es más probable que reconozcan el tipo de sitio que atraerá a sus compañeros, el tipo de problemas que les afectan a sus compañeros, y puedan hablar directamente de la experiencia sobre el tipo de discurso de odio o los sitios de odio que las personas de su comunidad suelen encontrar.

Nombre	**getthetrollsout**
Dirección web - contacto	https://www.getthetrollsout.org/
Quién la promueve	El Media Diversity Institute (MDI) trabaja internacionalmente para alentar y facilitar la cobertura responsable de la diversidad en los medios. Su objetivo es evitar que los medios de comunicación propaguen intencionadamente o involuntariamente los prejuicios, la intolerancia y el odio que pueden llevar a tensiones sociales, disputas y conflictos violentos. MDI alienta, en cambio, a una cobertura mediática justa, precisa, inclusiva y sensible para promover el entendimiento entre diferentes grupos y culturas. En partenariado con seis socios en Europa: Center for Independent Journalism (Hungría), Ligue Internationale Contre le Racisme et l'Antisémitisme (Francia), Symbiosis (Grecia), Amadeu Antonio Stiftung (Alemania), European Union of Jewish Students y European Forum of Muslim Women.
Breve descripción [1]	Este sitio web forma parte de un proyecto y una campaña para combatir la discriminación y la intolerancia por motivos religiosos en Europa. Dirigida por el Media Diversity Institute (MDI), el proyecto se basa en el poder de las redes sociales para difundir productos de medios innovadores y generar un diálogo para ofrecer una contra-narrativa poderosa contra diversas formas de discurso de odio, incluidos el antisemitismo, la islamofobia, el sentimiento anticristiano y los intentos asociados de cambiar la opinión pública contra los migrantes y los solicitantes de asilo (consulte la sección sobre odio antirreligioso para obtener más información). Objetivos del proyecto: • Reducir y degradar el discurso de odio, la discriminación y la intolerancia por motivos religiosos en los medios de comunicación europeos. • Capacitar a las organizaciones de la sociedad civil en Europa para identificar y destacar la intolerancia y la xenofobia dirigidas a grupos religiosos, incluidas las comunidades judías, musulmanas y cristianas. • Desafiar los estereotipos, desacreditando la mitología extremista y formando la opinión pública mediante el desarrollo de contenido inclusivo y aprovechando las plataformas de medios tradicionales y nuevos. • Consolidar y ampliar la red de defensores de los derechos humanos y jóvenes activistas comprometidos a frenar el aumento de la intolerancia y la discriminación en Europa.

Nombre	**getthetrollsout**
Breve descripción (2)	Actividades clave: • Monitoreo de medios: la base de la campaña será el monitoreo del discurso antirreligioso y del discurso del público y otras figuras en los medios tradicionales y nuevos. • Denuncias: cuando los esfuerzos de monitoreo del proyecto detecten discursos antirreligiosos en los medios de comunicación, los socios del proyecto los expondrán y los contrarrestarán utilizando los mecanismos más adecuados, como artículos, vídeos, publicaciones en blogs, cartas y reuniones con editores y jefes de política. Además de denunciar discursos de odio a plataformas de redes sociales. • Producción de vídeo: se producirán y compartirán vídeos en las redes sociales para desacreditar rigurosamente los estereotipos antirreligiosos y dar voz a testimonios y experiencias personales, contribuyendo a una mayor conciencia y comprensión del discurso de odio y su impacto. • Campaña de redes sociales: este sitio web y sus plataformas de redes sociales asociadas son el componente clave de la campaña y se utilizarán tanto para difundir los productos de la campaña como para para iniciar un debate sobre la lucha contra el odio antirreligioso. • Memes: la inmediatez y la ironía de los memes se utilizarán para contrarrestar los mensajes de odio que han continuado plagando este medio. Disponen de varias secciones como las siguientes: Vídeos. La sección proporciona una visión general de los resultados del monitoreo de los medios, desestima los estereotipos antirreligiosos y expresa testimonios personales y experiencias de discurso de odio. Monitoreo de medios. Esta sección contiene informes sobre el monitoreo de los medios de comunicación del discurso antirreligioso y el discurso del público y otras figuras en los medios tradicionales y nuevos. Quejas. Aquí se detallan las quejas de la campaña y los informes sobre el discurso de odio antirreligioso en los medios tradicionales y nuevos.

Nombre	**getthetrollsout**
Metodología	Empezaron la monitorización en los medios del odio antirreligioso en enero de 2018. En el primer año (2018) han detectado 310 incidentes. El Dr. Verica Rupar ha analizado los incidentes y ha hecho un informe disponible en pdf (ver casilla de enlaces). El principal criterio para la selección de los incidentes, (partiendo de palabras clave), ha sido la repercusión del incidente (la circulación, número de visitantes, número de visitas). Han analizado también la radio (no sólo medios escritos). Como medios escritos han buscado en periódicos online, Facebook y Twitter de los medios que han seguido y los post en las páginas web de los medios de comunicación. De los 310 incidentes analizados los de Islamofobia ocupan el primer lugar (45,61%) seguidos de discursos anti-inmigrantes o anti-refugiados (28,69%). El país donde más incidentes se han recogido ha sido en Bélgica (20,31%), seguido de Francia (13,75%) y Alemania (10%). En el informe hay un compendio de tópicos sobre los inmigrantes.
Trayectoria	Hay una segunda fase del proyecto: Bélgica, Francia, Alemania, Grecia, Hungría y Reino Unido (2017-2019).
Observaciones generales	Nombran el "ousted-troll of the month" y le hacen publicidad negativa: http://getthetrollsout.org/what-we-do/troll-of-the-month/item/353-alison-chabloz-ousted-troll-of-the-month.html
Links de interés	http://www.media-diversity.org/en/index.php?option=com_content&view=article&id=3229:get-the-trolls-out-phase-2&catid=18:current-projects&Itemid=21 Informe del primer año: http://getthetrollsout.org/resources/item/374-a-year-in-media-monitoring.html

Nombre	**Save a hater**
Dirección web - contacto	www.saveahater.accem.es
Quién la promueve	ACCEM en el marco de un proyecto de sensibilización denominado SiembraRED que tiene como objeto generar conciencia crítica entre la ciudadanía, promover la reflexión sobre el modo como utilizamos las redes sociales, luchar contra la discriminación y ayudar a evitar y atajar conductas potencialmente violentas.
Quién la financia	Financiado por el Ministerio de Sanidad, Servicios Sociales e Igualdad.
Breve descripción	En los últimos años se ha estado gestando un nuevo colectivo muy diferente pero que merece nuestra atención, las personas denominadas Haters. Personas que voluntariamente deciden quebrar la convivencia y provocar la descohesión social compartiendo y generando mensajes de odio hasta convertirse a sí mismas en un grupo de autoexclusión. Desde Accem se quiere combatir esta situación, sumando este colectivo a los que ya se les presta ayuda. Por eso lanzaron "SAVE A HATER", una iniciativa para intentar salvar a todos esos haters, ayudándoles a salir de su agujero y reintegrarlos en la sociedad. Estas personas, a su manera, también necesitan apoyo, como otros colectivos a los que atienden, pero para liberarles de su cerrazón y prejuicios.
Observaciones generales	Dispone de un manual para trabajar en contra-narrativa en diferentes ámbitos (polarización, ciberodio, misoginia en las redes, etc.).

Nombre	**Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online (HATE METER)**
Dirección web contacto	www.hatemeter.eu
Quién la promueve	Coordinador: eCrime, Universidad de Trento, Facultad de Derecho (IT) Socios: Fondazione Bruno Kessler Universite Toulose 1 Capitole Universidad de Teesside Amnistía Internacional Sezione Italiana Onlus Asociación de defensa de los derechos del hombre - Collectif contre l'Islamophobie en France Stop Hate
Quién la financia	REC Action Grant (REC-DISC-AG-2016) 24 months (01.02.2018 - 31.01.2020).
Breve descripción [1]	El proyecto Hatemeter trata de sistematizar, aumentar y compartir el conocimiento sobre el odio anti-musulmán en línea, y aumentar la eficiencia y efectividad de las ONG para prevenir y combatir la islamofobia a nivel de la UE, desarrollando y probando una herramienta de TIC (la plataforma HATEMETER) que monitorea y analiza automáticamente los datos de Internet y de las redes sociales sobre el fenómeno, y produce respuestas para apoyar las narrativas y las campañas de sensibilización. Hatemeter adopta un enfoque multidimensional para prevenir y combatir el odio anti-musulmán online en la UE, a través de un uso estratégico de Internet y las redes sociales, creando las condiciones para: • lograr una mejor comprensión del odio anti-musulmán online, explotando el potencial de Internet y las redes sociales; • entregar la plataforma Hatemeter a las ONG para ayudar a prevenir y combatir el odio anti-musulmán online; • fortalecer la cooperación entre los actores clave y garantizar una circulación más amplia y un impacto a largo plazo de los resultados del proyecto en futuras líneas de investigación y estrategias operativas.

ÍNDICE

Nombre	**Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online (HATE METER)**
Breve descripción [2]	Desde una perspectiva interdisciplinar (criminología, ciencias sociales, ciencias de la computación, estadísticas y leyes), Hatemeter analiza el problema en línea y evalúa las experiencias de fondo, necesidades y aspiraciones de las ONG y otros grupos de interesados directos, para redactar *Directrices sobre los requisitos sociotécnicos de la herramienta TIC.* La plataforma Hatemeter utiliza una combinación de procesamiento de lenguaje natural (NLP), aprendizaje automático y análisis / visualización de big data para: • Identificación en tiempo real: identificar y sistematizar en tiempo real las "banderas rojas" del discurso de odio anti-musulmán y/o posibles amenazas relacionadas en línea; • Comprensión en profundidad: comprender y evaluar los conjuntos de características y patrones asociados con las tendencias de la islamofobia en línea; • Respuesta táctica / estratégica: desarrollar una planificación táctica / estratégica efectiva contra el odio anti-musulmán en línea a través de la adopción del enfoque innovador de Persuasión Asistida por Computadora (CAP); • Producción de antinarrativas: producir un marco de contra-narrativa preciso para prevenir y combatir la islamofobia en línea, y crear campañas de sensibilización basadas en el conocimiento y a la medida. La plataforma Hatemeter se prueba con tres ONG de la UE donde la magnitud del problema es considerable pero no se han implementado respuestas sistemáticas (Francia, Italia y el Reino Unido), lo que permite a Hatemeter abordar varios objetivos: 1) Recomendación de la CE sobre medidas para combatir efectivamente el contenido ilegal en línea; 2) "Coloquio anual sobre los derechos fundamentales de la UE - Tolerancia y respeto: prevenir y combatir el odio antisemita y anti-musulmán en Europa"; 3) "Código de conducta para contrarrestar el discurso de odio ilegal en línea"; 4) "Agenda europea de seguridad". Con el fin de fortalecer la cooperación entre los actores clave y garantizar una circulación más amplia y un impacto a largo plazo de los resultados del proyecto en futuras líneas de investigación y estrategias operativas, el proyecto favorece el desarrollo de capacidades, la capacitación, la sostenibilidad y la transferibilidad de la plataforma Hatemeter entre otros grupos (p. ej., periodistas / medios de comunicación, funcionarios públicos).

Nombre	**Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online (HATE METER)**
Metodología [1]	Se rastrean eventos relevantes relacionados con la islamofobia, se analizan picos de temas específicos, se estudian cuentas emergentes y hashtags. Han publicado un entregable relacionado con los requisitos socio-técnicos sobre los que se implementan las tecnologías de rastreo y que son los siguientes: 1) Se centran en Twitter y Facebook, ya que se utilizan ampliamente para difundir el discurso islamofóbico con diferentes enfoques (el primero con ataques cortos y directos, el segundo con estrategias retóricas específicas y más sofisticadas). 2) Monitoreo de fuentes de noticias que están directamente vinculadas a sentimientos anti-musulmanes; por ejemplo The Daily Mail en el Reino Unido. 3) Seguimiento de perfiles específicos con presencia pública en torno a los cuales el discurso islamofóbico es más frecuente, por ejemplo, la Alianza Football Lad´s en el Reino Unido o políticos de derechas en Italia. 4) Siguiendo los hashtags, palabras clave y sus combinaciones recomendadas en las secciones anteriores para cada idioma de interés. En particular, vigilancia de la aparición simultánea de hashtags de políticos con palabras neutrales como "islamici" o "musulmani", que se encuentran con más frecuencia que el lenguaje ofensivo explícito. Desde un punto de vista técnico, el sistema se basa en las API de Twitter y Facebook, lo que permite acceder y analizar datos de perfiles públicos. Supervisan la información sobre usuarios, publicaciones y métricas de viralidad asociadas (por ejemplo, cuántos "me gusta", "retweets", comentarios, etc.). La información recopilada en la etapa anterior se analiza utilizando herramientas de procesamiento de texto para extraer la información más relevante relacionada con el odio anti-musulmán en línea, como los metadatos conectados a los mensajes (es decir, usuario, fecha, retweets, me gusta), la red en la que se disemina el discurso y el contexto textual relacionado; por ejemplo, temas relevantes, personas y lugares mencionados, hashtags concurrentes, etc. Para este fin, emplean el conjunto de aplicaciones basadas en Java Stanford CoreNLP166 para el procesamiento de textos, que admite todos los idiomas del proyecto y, por lo tanto, brinda un marco unificado. En cuanto a la extracción de palabras clave, se emplea la herramienta Keyphrase Digger (Moretti et al., 2015), ya que ha sido desarrollada por FBK y está disponible para italiano, inglés y francés, lo que permitiría sintonizar el algoritmo de ALRECO para cumplir los requisitos del proyecto para los tres idiomas.

Nombre	**Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online (HATE METER)**
Metodología [2]	Creación de una base de datos para la integración de datos estructurados / no estructurados. Toda la información extraída de las noticias en línea y las publicaciones de las redes sociales, y el contenido relacionado extraído de dichas fuentes, se almacenan en el "almacén de conocimientos" del proyecto para facilitar su actualización y recuperación. Se prevé implementar un repositorio con una estructura mixta, donde una base de datos relacional (MySql) estándar se conecte sin problemas con los archivos Json que provienen de las API de redes sociales, y que contienen toda la información sobre el contenido del mensaje y la red. Para cada mensaje o noticia recuperada con los módulos anteriores, tanto el resultado de la fase de procesamiento de texto, como los lugares y personas mencionados y las palabras, también se almacenan junto con la información del usuario, la fecha de emisión y otros metadatos. Para las publicaciones en redes sociales, el idioma del usuario definido en el perfil se almacena, de modo que, incluso si todos los mensajes están en la misma base de datos, es posible recuperar sobre la marcha análisis específicos de cada idioma. Por otro lado, también será posible realizar análisis comparativos en los tres idiomas, cuando se hayan utilizado algunos hashtags en los tres países de interés (por ejemplo, #MuslimInvasion). Se elige basar este análisis en el idioma asociado con un perfil de usuario, en lugar de utilizar herramientas de detección de idioma, ya que estos algoritmos generalmente no funcionan bien en textos cortos. Dado que las publicaciones también pueden contener enlaces a fuentes de noticias, que generalmente describen un evento que provoca un comentario o una discusión, la información relacionada con las noticias también se almacena. Esto es importante para estudiar las fuentes típicas de información (errónea) discutidas en línea o para proporcionar pruebas para verificar los hechos.
Observaciones generales	Iniciativa interesante en el seguimiento y análisis de los delitos de odio por motivos religiosos.

Nombre	**Silence Hate**
Dirección web - contacto	www.silencehate.it
Quién la promueve	Tres organizaciones de voluntariado: Zaffiria, COSPE y Priscilla.
Quién la financia	UNAR. Oficina Nacional contra la Discriminación en Italia.
Breve descripción	El proyecto #SilenceHate - Digital Youth against Racism tiene como objetivo combatir la difusión del discurso de odio contra los migrantes y las minorías en Internet, a través de la educación sobre medios de comunicación de los jóvenes. A través de sus actividades, pretende proporcionar a los docentes, educadores y jóvenes, herramientas analíticas y operativas para reconocer y combatir el discurso de odio en línea, difundir el valor positivo de la diversidad y fomentar una cultura de respeto. La educación y la sensibilización son las estrategias más efectivas para combatir y prevenir el discurso de odio en Internet, y la escuela está a la vanguardia de la difícil tarea de enfrentar este fenómeno. Objetivos principales: • Combatir la difusión del discurso de odio en Internet hacia los inmigrantes y las minorías a través de la educación en medios para jóvenes. • Promover el papel activo de los jóvenes en la lucha y prevención del racismo en línea y el discurso xenófobo. • Sensibilizar a los jóvenes y a la opinión pública sobre el discurso de odio y sobre los riesgos de la proliferación descontrolada de mensajes racistas y xenófobos en sitios web y redes sociales. Las actividades planificadas comenzarán con la capacitación en cada territorio, dirigida a maestros y educadores, para encontrar ideas sobre como abordar el discurso de odio con sus estudiantes, a través de la educación en medios, el enfoque intercultural y la participación activa de niños y niñas. Posteriormente, se realizan talleres educativos en escuelas secundarias y centros juveniles, y se crea un sitio web y un módulo de capacitación para difundir metodologías, herramientas y resultados mediáticos de los cursos. Un evento público que involucra a interlocutores clave a nivel institucional y que tiene como objetivo promover una reflexión más amplia sobre los peligros de la propagación de la xenofobia y el racismo en la web y las posibles estrategias de contraste legal, tecnológico y cultural.

Nombre	**Silence Hate**
Trayectoria	Marzo 2018 – Marzo 2019.
A quién se dirige	Jóvenes, profesores, educadores, operadores y opinión pública. Emilia Romagna, Veneto, Toscana e Campania.

Nombre	**Contro l'odio**
Dirección web contacto	www.controlodio.it Marcos Stranisci es el responsable del proyecto desde la Asociación Acmos.
Quién la promueve	"Contra el odio" es un proyecto de Acmos, organización sin ánimo de lucro que desde 1999 se ocupa de la educación para la ciudadanía en las escuelas. Con la difusión de la web 2.0, Acmos ha comenzado a planificar iniciativas en línea para crear conciencia sobre temas de ciudadanía. Uno de estos es contra el odio.
Quién la financia	Financiado por el Ministerio de Trabajo y Políticas Sociales italiano.
Breve descripción	Desarrolla tecnologías informáticas para encontrar automáticamente discursos de odio publicados en línea. Concebido como un intento de responder a los problemas actuales relacionados con la presencia del odio en la web, el proyecto tiene tres objetivos principales: 1. Encontrar declaraciones de odio en línea mediante la creación de herramientas informáticas para la detección automática. 2. Sensibilizar a la ciudadanía, y en particular a los jóvenes, sobre la importancia de la comunicación responsable a través de la creación de laboratorios en diferentes territorios de nuestro país. 3. Contar los proyectos y las realidades que en Italia promueven una cultura de tolerancia. El proyecto contempla la construcción de tres herramientas: • El mapa del odio. Una serie de visualizaciones interactivas que muestran la cantidad de discursos de odio publicados en Twitter día a día. • El detector de odio. Una herramienta que le permite al usuario analizar la cantidad de odio presente en su red social. • El mapa de tolerancia, donde se recogen y muestran todos los proyectos y las realidades que favorecen la inclusión social.
Observaciones generales	Disponen de un algoritmo interesante.

Nombre	**Hate Base**
Dirección web contacto	www.hatebase.org
Quién la promueve	Hate Base
Breve descripción	Hatebase es una plataforma de software diseñada para ayudar a las organizaciones y comunidades online a detectar, monitorear y poner en cuarentena el discurso de odio. Los algoritmos de Hatbase analizan conversaciones públicas utilizando un amplio vocabulario basado en la nacionalidad, etnia, religión, género, orientación sexual, discapacidad y clase, con datos en más de 80 idiomas y más de 200 países. El discurso de odio degrada la conversación pública, silencia los diversos puntos de vista, y puede ser un indicador de alerta temprana de la violencia. Hatebase es una empresa con sede en Toronto cofundada por Timothy Quinn y The Sentinel Project, una organización canadiense sin fines de lucro dedicada a reducir las atrocidades masivas en países como Kenia, Myanmar y la República Democrática del Congo. La misión de Hatebase es reducir los incidentes de incitación al odio al monitorear el uso y la difusión del lenguaje discriminatorio contra grupos específicos, disminuir la aceptabilidad del discurso de odio, y prevenir la violencia que es predicada por el discurso de odio. Hatebase usa un amplio vocabulario multilingüe basado en nacionalidad, etnia, religión, género, discriminación sexual, discapacidad y clase para monitorear incidentes de odio en más de 200 países. Nuestro motor de lenguaje natural, Hatebrain, realiza análisis lingüísticos en conversaciones públicas para derivar una probabilidad de contexto odioso. Todos los datos están disponibles a través de la interfaz web y la API de Hatebase. Nuestro vocabulario regionalizado y datos de avistamientos son útiles para monitorear las tendencias en el uso del odio y para correlacionar con otros conjuntos de datos con el objetivo de realizar análisis de riesgo de conflictos.
A quién se dirige	Como empresa, se dirige a un amplio espectro de potenciales clientes. Parece que existe una licencia gratuita para ONG, para organismos gubernamentales y para el ámbito académico, aunque las consultas diarias están restringidas a 100.

Nombre	**Donate the hate**
Breve descripción	Internet está actualmente saturado de comentarios racistas y xenófobos. DONATE THE HATE - la primera iniciativa de caridad en línea involuntaria. La idea detrás de esto es la siguiente: por cada comentario de odio, hacemos una DONACIÓN DE 1 EURO a los proyectos de refugiados dirigidos por la campaña "Aktion Deutschland Hilft" y "EXIT-Deutschland", una iniciativa contra el extremismo de derecha. De esta manera, los que odian y los trolls están haciendo una donación contra su propia causa: proporcionan los fondos que utilizan en la campaña para convertir los comentarios de odio en donaciones involuntarias. Para la implementación, se ha diseñado un micrositio y se ha creado una página de Facebook, que es el centro. Con la página de Facebook, los usuarios registrados pueden convertir semiautomáticamente, a través de una interfaz, comentarios de odio en donaciones involuntarias. Para este propósito, se ha diseñado y programado una APP, que hace posible el comentario a través de una interfaz API. En el micrositio se pueden encontrar algunos comentarios de odio comentados, así como uno de los diez principales donantes involuntarios. Todo anónimo. Donaciones para la acción proveniente de socios como: emisoras de radio, periódicos y estaciones de televisión. Facebook también apoya la campaña.

Nombre	**Exploring Online Hate**
Dirección web - contacto	https://exploringhate.newamerica.org/
Quién la promueve	New America es pionera en un nuevo tipo de pensamiento y acción: una plataforma cívica que conecta a distintos organismos (un instituto de investigación, un laboratorio de tecnología, una red de soluciones, un centro de medios y un foro público). Diseñamos y promovemos políticas públicas basadas en la evidencia. Una de sus líneas de actuación es desarrollar herramientas legales, políticas y tecnológicas para desarrollar la capacidad democrática. ADL y el Centro para la Tecnología y la Sociedad: ADL es una organización anti-odio fundada hace 105 años para proteger al pueblo judío. Su objetivo es luchar contra el incremento de todas las formas de odio y garantizar justicia y un trato justo para todos, especialmente la comunidad judía, los inmigrantes, los miembros de minorías raciales, étnicas y religiosas y la comunidad LGBTQ. Otra de sus líneas de actuación es la formación, ADL capacita a 1.5 millones de maestros y escolares cada año en programas educativos contra el prejuicio y lidera la construcción de coaliciones de múltiples credos y derechos civiles.
Quién la financia	Sin datos.

Nombre	**Exploring Online Hate**
Breve descripción	Ante el incremento del odio en redes sociales e internet han diseñado una herramienta interactiva para monitorear y estudiar la actividad extremista en Twitter, basada en poco más de 1.000 cuentas que utilizan regularmente contenido de odio contra grupos protegidos. El enfoque de la herramienta proporciona una comprensión visual y cualitativa de la escala y el alcance de la actividad de odio en línea casi en tiempo real. El proyecto monitorea el contenido, los temas y la actividad de más de 1.000 cuentas en Twitter. La construcción del conjunto de cuentas rastreadas comenzó con la identificación de 40 "cuentas semilla" que participan regularmente en la difusión de contenido de odio contra grupos vulnerables. Las cuentas semillas fueron identificadas por un grupo de expertos independientes. Posteriormente, utilizan las API públicas de Twitter para generar una lista de seguidores para cada una de las cuentas semilla, y a continuación agrupan todas esas "cuentas seguidoras" en un conjunto más amplio, clasificándolas por el número de cuentas semilla que siguen. A través de una combinación de métodos algorítmicos, generan una lista de las principales cuentas que participan en una conducta de odio entre estas cuentas seguidoras. El conjunto de cuentas resultante se rastrea en este panel e incluye más de 1.000 cuentas, aunque el número exacto de cuentas puede variar debido a las suspensiones y eliminaciones que administra Twitter. Finalmente, utilizan una combinación de API de propiedad pública y disponible para analizar los datos asociados con las cuentas de seguidores en tiempo real. Este análisis potencia las visualizaciones de datos disponibles en el panel de control.
Trayectoria	Este panel fue concebido tras el incidente en Charlottesville, Virginia, en agosto de 2017. Durante el año siguiente, el panel se desarrolló en New America con el apoyo y la guía de la Liga Anti-Difamación. El panel busca no solo explorar y comprender mejor la dinámica social y tecnológica de la actividad de odio en Twitter, sino también generar y dar forma a un amplio debate público sobre el odio en línea.
Observaciones Generales	Se va a indagar más sobre esta experiencia en una segunda fase de trabajo.

Nombre	**SaferLab**
Dirección web - contacto	http://saferlab.org.br/
Quién la promueve	SaferNet, organización que trabaja desde 2005 en la promoción de derechos humanos en la web, se ha unido a Google.org y a UNICEF para proponer un nuevo enfoque relacionado con el discurso de odio en la red.
Quién la financia	Google.org
Breve descripción	SaferLab es un laboratorio creativo que tiene como objetivo inspirar, potenciar y apoyar el protagonismo de los jóvenes en la producción de narrativas para abordar el discurso de odio y la discriminación en línea basada en el género y la etnicidad, dentro de los principios de los Derechos Humanos. Los objetivos del laboratorio creativos son los siguientes: 1. Inspirar a los jóvenes, dándoles contenido para compartir, enfocándose en narrativas positivas, datos e información con un énfasis en los Derechos Humanos y la diversidad para promover su empatía. 2. Capacitar a los participantes proporcionándoles herramientas para un enfoque innovador de contra-narrativa, como técnicas de narrativa y estrategias de comunicación en línea. 3. Apoyarlos para implementar sus proyectos con mini subvenciones, desarrollo de capacidades, mentores y recursos.

Nombre	**SaferLab**
A quién se dirige	SaferLab se organiza en una serie de fases de trabajo que se resumen a continuación: 1. Inscripciones y selección. Pueden participar jóvenes de entre 16 y 25 años interesados en temas relacionados con los Derechos Humanos. 2. Resultados de la primera etapa. Selección de 300 personas para la primera fase de formación con encuentros en línea. 3. Encuentros en línea. Actividades semanales en línea sobre temas relacionados con los Derechos Humanos y la gobernanza de Internet. 4. Resultado de la segunda etapa. Selección de las 150 personas más comprometidas. 5. En la carretera. Fines de semana de inmersión presencial, vivencial y con producción de contenidos. 6. Mentorías con profesionales para ayudar a sacar las ideas del papel. 7. Financiación. Selección de los diez mejores proyectos para las becas entre R$ 1,5 mil y R$ 12 mil.
Puntos positivos	Junto a estas fases de trabajo con los jóvenes, SaferLAb dispone también de otras líneas de actuación: • Elaboración de materiales. Disponen de una guía muy visual y sencilla de utilizar para la creación de contra – narrativas. • Reciben denuncias. Desde 2006 la organización ha recibido 2.061.141 denuncias por delitos de odio de los cuales casi un tercio (28%) eran por racismo.

Nombre	**Observatorio de la Islamofobia en los medios**
Dirección web - contacto	http://www.fundacionalfanar.org/ www.observatorioislamofobia.org
Quién la promueve	Fundación Al Fanar.
Quién la financia	IEMed, Fundación Al Fanar, La Fundación Tres Culturas, Casa Árabe y la Fundación Euroárabe.
Breve descripción	El Observatorio parte de la premisa de que los medios de comunicación tienen un papel principal a la hora tanto de forjar como de romper estereotipos e imágenes distorsionadas o discriminatorias, en el contexto actual de auge de expresiones racistas y discriminatorias dirigidas a personas de confesión musulmana. Entre enero y diciembre de 2017 el equipo el Observatorio hizo un seguimiento de seis diarios nacionales (La Vanguardia, El Mundo, La Razón, 20 Minutos, El País y Diario.es) para identificar informaciones islamófobas y emitir recomendaciones. En la página del Observatorio se puede acceder a los artículos creados para el Observatorio, a los ejemplos de artículos islamófobos analizados, a los ejemplos de buenas prácticas, y a un glosario de términos sobre el islam. El sitio ofrece además, diferentes recursos de material de apoyo y una agenda con eventos relacionados con la lucha contra este fenómeno. Se trata de un espacio interactivo donde los usuarios pueden colaborar denunciando o haciendo sus aportaciones al proyecto.
Trayectoria	Incio en 2017.
Observaciones generales	Islamofobia.

Nombre	**INFO RAXEN- Servicio de Noticias de Movimiento contra la Intolerancia**
Dirección web - contacto	http://www.informeraxen.es/
Quién la promueve	Movimiento contra la Intolerancia.
Breve descripción	El INFORME RAXEN (Racismo, xenofobia, antisemitismo, islamofobia, neofascismo, homofobia e intolerancia a través de los hechos) recoge noticias y situaciones de discriminación, agresiones y violencia, geografía de conflicto, seguimiento de publicaciones racistas, manifestaciones xenófobas en el ámbito de la cultura, deporte y sociedad, en Internet, conflictos escolares… y en todos aquellos ámbitos que tienen que ver con la intolerancia y el racismo, en una búsqueda de datos en la que participa el tejido asociativo del Movimiento contra la Intolerancia. Este Informe responde a una necesidad de acercarnos a la realidad de los hechos de racismo, xenofobia, antisemitismo, islamofobia, neofascismo, homofobia y otras manifestaciones de intolerancia, permitiendo conocer en proximidad y alcance territorial unos datos muy útiles para profesores, asociaciones, políticos y personas preocupadas por esta problemática. Responde a su vez a diversos mandatos de las Instituciones Europeas sobre la publicación de datos sobre delitos de odio y discurso de intolerancia, y persigue la sensibilización de la sociedad de acogida e instituciones, en defensa de los valores democráticos. Los Cuadernos de Análisis recogen estudios, informes, documentos internacionales, analizan normativas y campañas, y difunden un discurso de vanguardia e innovador en la lucha contra el racismo y la intolerancia, así como en valores y en defensa de las víctimas de los Crímenes de Odio. Muy valorados entre instituciones, educadores, ONG, centros de documentación, Facultades de Derecho, Trabajo Social, Educación Social e investigadores y personas interesadas. Los Materiales Didácticos, compilan cuestiones temáticas y están orientados a facilitar documentación de interés para animadores culturales, profesores, concejales, periodistas y personas solidarias de forma que sean instrumentos que les permitan abordar con solvencia la temática en cuestión. Para su realización se lleva a cabo un proceso de búsqueda de documentación, estudio, contraste y análisis.

Nombre	**INFO RAXEN- Servicio de Noticias de Movimiento contra la Intolerancia**
Trayectoria	El 22 de enero de 1994 se lanza el primer informe RAXEN. A partir del año 2000 hasta el año 2012 el Informe RAXEN se elaboraba con periodicidad trimestral. A partir del año 2013 hasta la actualidad el Informe RAXEN se elabora con periodicidad mensual.
A quién se dirige	Administraciones Públicas, ONG, centros de documentación, Facultades de Derecho, Trabajo Social, Educación Social e investigadores y personas interesadas.
Observaciones generales	Una larga trayectoria que permite tener una visión longitudinal.

INFORME 6

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

Conclusiones y recomendacines

ÍNDICE

6. Conclusiones y recomendaciones

EN BASE A LAS EXPERIENCIAS Y BUENAS PRÁCTICAS ANALIZADAS SE RECOMIENDA/CONCLUYE:

Salvo un par de excepciones, la mayoría de las experiencias analizadas no son herramientas que identifiquen o monitoricen el discurso de odio en internet si no que combinan diferentes acciones como la detección, la sensibilización, la contranarrativa, la formación, etc.	El idioma en el que se plantea la herramienta es la clave, más allá de la ubicación geográfica de la experiencia. Si se quiere llegar a determinada población es necesario pensar en una herramienta multilingüe (árabe, francés, inglés, español, etc.).	La perspectiva de género no se incluye en la mayoría de las experiencias y sería recomendable que ALRECO la tuviera presente en el diseño de su herramienta (múltiples discriminaciones de las mujeres).	No se ha podido indagar lo suficiente en cada una de las experiencias. Es necesario tomar este informe como un punto de partida para el resto de las tareas previstas en el proyecto ALRECO y, esencialmente, para el diseño del sistema de indicadores.
Es imprescindible contar con un objetivo acotado y delimitado para el desarrollo de la herramienta prevista en el proyecto ALRECO. EL tiempo de ejecución del proyecto es de 24 meses y para el desarrollo de la herramienta, testeo y validación de esta en base a las experiencias analizadas, los plazo justos.	La mayoría de las experiencias se realizan en Twitter ya que Facebook no parece técnicamente accesible.	Es recomendable realizar un análisis de concordancia de los validadores para identificar sesgos o discrepancias, así como realizar algún tipo de monitoreo y seguimiento de los mismos de cara a introducir mejoras en la herramienta.	Se han identificado expertos con los que sería recomendable contactar en nombre del proyecto e incluirles en alguna fase del proyecto como grupo experto, conferencia final, etc.

ÍNDICE

INFORME 7

BUENAS PRÁCTICAS Y EXPERIENCIAS
en la UE para analizar el
discurso de odio en línea

ANEXOS

7.1. Revisión literatura académica

7.2. Webgrafía

ÍNDICE

1. INFORMACIÓN GENERAL

Mejora de un clasificador de discurso de odio en español y desarrollo de una herramienta que analiza el estado de odio en Twitter. Juan Carlos Pereira Kohatsu (2018)
Tutor: Lara Quijano Sánchez Co-tutores: Álvaro Ortigosa Juárez, Miguel Camacho Collado

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Los delitos de odio son un tipo de violación de la ley cuya motivación primaria es la existencia de prejuicios respecto a las víctimas de este delito, y que tienen lugar cuando el infractor elige a sus víctimas en terrenos que pertenecen a cierto grupo definido básicamente por cualidades, como la etnia o la raza, nacionalidad, idioma, orientación sexual, religión, discapacidad física o identidad de género, entre otros. El discurso de odio consiste en la manifestación de expresiones para alentar el odio y la discriminación por las cualidades mencionadas antes. Hay evidencia que estos delitos están influenciados por eventos ampliamente publicitados (actos terroristas, migración, manifestaciones, revueltas…). Este tipo de eventos generalmente actúan como desencadenantes, lo que aumenta dramáticamente la frecuencia de este tipo de delitos. La principal dificultad para estudiar el discurso de odio en redes sociales, una vez que el problema de la recopilación de datos está resuelta, es la correcta clasificación de "tweets" en dos categorías (odio y no-odio) usando diferentes técnicas como el "Natural Language Processing" (Procesamiento Natural del Lenguaje) para extraer atributos y patrones de los textos para finalmente clasificar los mensajes de las herramientas de Machine Learning como odio o no odio. El elemento clave en el proceso de clasificación es el contenido del mensaje en el que el autor del "tweet" expresa sentimientos u opiniones sobre una entidad o grupo de personas. Actualmente, la Secretaría de Estado de Seguridad de España, y más concretamente la Oficina de Delitos de Odio, está interesada en realizar análisis de mensajes en Twitter y otras redes sociales, para detectar discursos de odio, reaccionar ante ellos y mejorar la seguridad y el bienestar de la sociedad española. Dada la creciente necesidad de monitorear los mensajes de odio en redes sociales que pueden ser potencialmente peligrosos, la meta principal de esta tesis es renovar y mejorar el rendimiento de los clasificadores del discurso de odio y desarrollar una herramienta de análisis de datos que identifique los tweets nocivos y otras construcciones de sentimientos negativos.

METODOLOGÍA
El modelo fue construido siguiendo estos pasos: Recopilación de 885.758 tweets que fueron descargados para construir un corpus para el modelo de entrenamiento. Estos tweets fueron obtenidos usando Twitter Rest API durante un corto periodo de tiempo. Las restricciones para descargar tweets fueron: que los tweets debían estar escritos en español y proceder de España. Para responder a esta limitación fueron creadas 12 aplicaciones de Twitter y rotadas cuando la anterior excedió el número de solicitudes (20 solicitudes por 15 min). El contenido de la limpieza de datos es un texto en formato unicode, así que pueden contener: emojis, tildes, chinos y otros caracteres no ASCIE. Para simplificar el problema, todos los caracteres no alfanuméricos fueron eliminados. Durante esta etapa se crearon algunos tokens para unificar diferentes expresiones.

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS

IMPACTO (ámbito de influencia de la iniciativa)
Herramienta empleada por la Oficina de Delitos de Odio de la Secretaría de Estado de Seguridad de España. Ministerio del Interior.

2. INFORMACIÓN GENERAL

Construcción de modelos de clasificación automática para discursos de Odio, Universidad Auntónoma de Madrid, Juan Carlos Pereira Kohatsu (2017)
Tutor: Alvaro Ortigosa
https://repositorio.uam.es/handle/10486/680053

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS

Las llamadas redes sociales constituidas por plataformas tales como Facebook™, Twitter™ constituyen el soporte de los medios de comunicación sociales que facilitan el intercambio y la discusión de información, experiencias y opiniones entre individuos de manera rápida y masiva. La explosión de los medios sociales ha tenido consecuencias que han sido valoradas tanto positiva como negativamente para el conjunto de la sociedad. Entre los efectos negativos, los medios sociales han hecho 'visibles' algunas actitudes de ciertos grupos sociales que se traducen en ataques a personas o colectivos en razón de su pertenencia a unos determinados grupos definidos por características de nacionalidad, preferencias sexuales, raza, religión, que, en muchos países, han sido catalogados como delitos de odio (1). Así pues, nace la necesidad de desarrollar un sistema que permita determinar si el autor de un mensaje es perpetrador de delitos de odio o no en una determinada red social, tarea nada sencilla de realizar puesto que la inmensa mayoría de los mensajes en las redes sociales no son de odio. Este proyecto toma como referencia la red social Twitter y los distintos tweets (microblogs) generados en la misma. Para crear un sistema de detección de tweets de odio en primer lugar será necesario desarrollar un modelo predictivo y posteriormente encapsularlo en un clasificador para que pueda utilizarse por el usuario final. Durante el desarrollo del modelo ha sido necesaria la descarga de una gran cantidad de tweets haciendo uso de la API de Twitter y la posterior limpieza de estos, reduciendo la cantidad de ruido existente en los propios mensajes como repetición de caracteres y uso de símbolos extraños tales como los emojis. Además, se ha hecho uso de técnicas de procesamiento de lenguaje natural (NLP) que han permitido extraer información de los tweets previamente procesados. Entre otras herramientas, ha sido necesario entrenar un analizador morfológico (POS-Tagger) para extraer clases de palabras que concentren la mayor parte de la semántica del mensaje (verbos, nombres y adjetivos). Tras el tratamiento de los tweets, se ha desarrollado un filtro que permite equilibrar la cardinalidad de ambas clases de mensajes (odio y no odio) pasando de una proporción de 1:1000 a 270:1000. Una vez realizado el sobremuestreo de tweets con contenido de odio, se procede a la aplicación de técnicas de clasificación supervisada basadas en aprendizaje máquina, siendo las redes neuronales profundas el mejor clasificador para enfrentar este enmarañado problema. Finalmente, tras la validación, se ha creado un clasificador, basado en el modelo desarrollado, que permite al usuario detectar al vuelo los tweets de odio.

3. INFORMACIÓN GENERAL

Aplicar una detección automática basada en el texto sobre el lenguaje engañoso para las denuncias a la policía: extracción de patrones de comportamiento de un modelo clasificativo en distintas fases para entender cómo se intenta engañar a la policía (2018)
Lara Quijano-Sánchez, Federico Liberatore, José Camacho-Collados, Miguel Camacho-Collados
doi: 10.1016/j.knosys.2018.03.010

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Redactar una denuncia a la policía falsa es un delito que tiene serias consecuencias tanto para el sistema como para el individuo. De hecho, puede suponer una acusación tanto de delito menor como mayor. Para la sociedad, un informe falso resulta en la pérdida de recursos policiales y la contaminación de sus bases de datos de investigaciones y evaluaciones sobre el riesgo de crimen de un territorio. En esta investigación, presentamos VeriPol, un modelo para la detección de denuncias falsas a la policía basadas únicamente en el contenido de su texto. Esta herramienta, desarrollada en colaboración con la Policía Nacional de España, combina los métodos de Procesamiento de Lenguaje Natural y Aprendizaje de Máquina (en inglés, Natural Language Processing and Machine Learning) en un sistema de apoyo a las decisiones que provee a los oficiales de policía la probabilidad de que una denuncia sea falsa. VeriPol ha sido testeado en más de 1000 informes desde el año 2015 provenientes de la Policía Nacional Española. Los resultados empíricos demuestran que es extremadamente efectiva en realizar una discriminación entre las denuncias verdaderas y falsas, con un margen de éxito de más del 91%, mejorando en más de un 15% la precisión de la policía en el mismo set de datos. El modelo de clasificación subyacente puede ser analizado para extraer patrones y perspectivas que muestren cómo hay personas que tratan de engañar a la policía (incluyendo el hecho de realizar denuncias a la policía falsas sin ser detectados). En general, a mayor sea el número de detalles presentados en el informe, mayor es la probabilidad de que sea fiable. Finalmente, un estudio piloto realizado en junio del año 2017 ha demostrado la utilidad de VeriPol en su campo de trabajo.

DESARROLLO Y ALCANCE DE LA INICIATIVA

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS
- VeriPol es un modelo efectivo de detección de mentiras en las denuncias a la policía basándose en el texto.
- Nuestro modelo incluye selección de características en base a reglas heurísticas y penalización L1.
- Los experimentos computacionales en un set de datos real muestran una precisión de validez del 91%.
- Un estudio piloto muestra un límite menor de precisión empírica de un 83%, aproximadamente.
- El análisis modelo provee percepciones lingüísticas acerca de cómo hay personas que tratan de engañar a la policía.

4. INFORMACIÓN GENERAL

Cómo los modificadores de género y de tonalidad de piel afectan a la semántica de Emojis en Twitter
Franceso Barbieri (Universidad Pompeu Frabra, Barcelona, Spain), Jose Camacho Collados (Cardiff Universidad, Reino Unido)
http://www.aclweb.org/anthology/S18-2011

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
En este artículo se analiza el uso de emojis en las redes sociales en relación con la tonalidad de piel y el género. Mediante la recolección de set de datos de más de 22 millones de tweets provenientes de los Estados Unidos, se pueden obtener algunos resultados después de realizar un análisis simple a partir de las frecuencias. Además, se ha llevado a cabo un análisis semántico sobre el uso de emojis y sus variedades (ej. Variedad de género y de color de piel) mediante la incrustación de todas las palabras, emojis y variedades en el mismo espacio vectorial. Los análisis revelan que algunos estereotipos relacionados con el color de piel y el género se reflejan en el uso de estas variantes. Por ejemplo, los emojis que representan gestos de manos se utilizan más comunmente con tonos de piel más claros, y su uso con distintos tonos de piel varía significativamente. De la misma manera, el vector correspondiente a la variante masculina tiende a ser semánticamente más cercano a los emojis relacionados con los negocios o la tecnología, mientras que la variante femenina aparece más relacionada con emojis de amor o maquillaje.

METODOLOGÍA
Se planteó el problema desde dos perspectivas metodológicas. Primero, se analizó el uso de emojis y sus variantes desde un punto de vista numérico, contando sus apariciones en un corpus. Esto ya proporcionó una visión importante sobre cómo estos emojis son usados. Luego, se combinó con el modelo incrustado SW2V (Sentidos y Palabras para los Vectores o *Senses and Words to Vectors* en inglés) (Mancini et al., 2017) para entrenar a un vector especial en el que los emojis y sus variedades se codifiquen juntos, permitiendo analizar su interpretación semántica.

5. INFORMACIÓN GENERAL

Series de Investigación de los Derechos Humanos – Instituto Holandés de Derechos Humanos (SIM)
https://intersentia.com/en/product/series/show/id/9166/
Marloes van Noorloos (Universidad de Tilburg) y Antoine Buyse (Universidad de Utrecht).

6. INFORMACIÓN GENERAL

Algoritmos para combatir el discurso de odio online, Dr. Uwe Bretschneider (2017)
https://www.research-in-germany.org/en/infoservice/newsletter/newsletter-2017/october-2017/algorithms-to-combat-hate-speech-online.html

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
"El programa analiza los comentarios y busca palabras y grupos de palabras almacenadas en su base de datos", según Bretschneider. Lo que es único de este software es su habilidad para reconocer el contexto en el que el insulto es usado. Como esas palabras individuales están unidas a personas o grupos, se describe en un algoritmo en el que se permite que el software también detecte la persona o grupo de personas al que el abuso es dirigido. "No es una cuestión de censura, sino un entendimiento sobre cómo se expresan las opiniones".

7. INFORMACIÓN GENERAL

NORIEGA, Chon A. e IRIBARREN, Javier (2009): "Discurso de odio en anuncios de radio. Informe preliminar sobre un estudio piloto". Latino Policy & Issues Brief, UCLA, 22: http://www.chicano.ucla.edu/press/briefs/documents/PB22_000.pdf
Este proyecto de investigación es una colaboración entre el UCLA Chicano Studies Research Center y la Coalición Nacional Hispánica de los Medios de Comunicación. Se apoya en parte de una subvención *del Programa de Conocimiento Necesario para una Esfera Pública Democrática* del Consejo de Investigación de Ciencias Sociales, con fondos otorgados por la Fundación Ford.

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Analizar el discurso de odio en anuncios comerciales de radio para identificar objetivos y tipos de discurso.

METODOLOGÍA
En octubre de 2008 el Buró Federal de Investigaciones (FBI) publicó su estadística anual de discurso de odio más reciente. Se eligió la radio para el estudio porque tiene la mayor tasa de penetración de entre los medios de comunicación de cualquier tipo de salida (impreso, digital o difundido), alcanzando al 90% de los americanos cada semana. Tres programas fueron seleccionados para el estudio piloto: El Show de Lou Dobbs dirigido por Lou Dobbs, La Nación Salvaje de Michael Savage, y el Show de John & Ken dirigido por John Kobylt y Kenneth Chiampu. Cada uno representa una franja de distinto tipo de anuncios comerciales.
Se examinó la transcripción de un segmento ininterrumpido de un cuarto del total de cada uno de los tres programas mencionados; todos ellos fueron transmitidos en julio de 2008 utilizando en la definición de NTIA sobre el discurso de odio (Departamento de Comercio de los Estados Unidos, año 1993).

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS. RESULTADOS
(1)"palabras que amenacen o inciten 'una acción ilegal inminente', que puede ser criminalizada sin necesidad de violar la Primera Enmienda"; o (2) "el discurso que crea un clima de odio o prejuicio, que al final pueda respaldar el acto de crímenes de odio". (Departamento de Comercio de los Estados Unidos, año 1993).
Los resultados se dividen en dos áreas: los objetivos a los que va dirigido el discurso de odio y los tipos de discurso de odio.
Se identificaron y desarrollaron seis categorías de grupo a los que va dirigido el discurso de odio que crea "un clima de odio y prejuicio." Tres de las seis categorías representan a grupos vulnerables: extranjeros, minorías étnicas y raciales, e individuos e instituciones identificadas con una creencia religiosa.
Las otras tres representan instituciones sociales vistas como cómplices de estos grupos vulnerables: organizaciones políticas, medios de comunicación, y el sistema de justicia criminal.
Identificamos cuatro tipos de discurso que, mediante declaraciones negativas, crean un clima de odio y prejuicio: (1) datos falsos, (2) argumentación defectuosa, (3) lenguaje divisorio (del estilo de "nosotros contra ellos"), (4) metáforas deshumanizantes.
La información muestra un patrón de retórica recurrente en el que los grupos vulnerables eran identificados como opuestos a los valores que el hacedor atribuye a la sociedad. Estos grupos luego eran relacionados con las instituciones sociales presentadas como cómplices. En efecto, los grupos a los que iba dirigido el discurso de odio eran caracterizados como una amenaza directa al estilo de vida de los oyentes.el

IMPACTO (ámbito de influencia de la iniciativa)

Los usuarios académicos se beneficiarán de dos maneras: primero, los investigadores tendrán un mayor entendimiento de la naturaleza, la manifestación y la difusión de información de odio online. Segundo, una vez validado el modelo probabilístico se puede adaptar a las alternativas de examen e interrogación por las ciencias sociales en áreas como la salud, la economía, los estudios de los medios de comunicación y la educación. Esta herramienta permitirá a los usuarios no académicos por otra parte incrementar su entendimiento de este fenómeno creciente y predecir la difusión de contenido de odio en redes digitales, otorgando así una oportunidad para su intervención antes de que este contenido se viralice, causando daño a los individuos, grupos minoritarios y comunidades.

8. INFORMACIÓN GENERAL

Comprender el discurso de odio antiinmigrante en las redes sociales
http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8952399&fileOId=8952403

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Identificar problemas centrales que han contribuido al discurso de odio hacia los inmigrantes. Abordar el discurso de odio en las redes sociales es una tarea difícil. Implica tres conjuntos de expresión que necesitan ser considerados para poder poner restricciones en las redes sociales: derechos de los individuos que expresan su opinión, derechos de las plataformas de redes sociales y las terceras partes. (…) Además, el derecho a la igualdad de trato de aquellos que son víctimas de este tipo de discursos, también necesita tomarse en cuenta. El objetivo del discurso de odio es ridiculizar a las víctimas, para humillarlas y representar sus agravios como menos serios. El discurso de odio crea una condición previa para un crimen de odio, ya que los crímenes de odio tienen más probabilidades de ocurrir con la estigmatización previa y la deshumanización de las víctimas seleccionadas.

METODOLOGÍA
Enfoque principal en el Pacto Internacional de Derechos Civiles y Políticos (PIDCP) y la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial (CERD). Con respecto a los documentos regionales, el Convenio Europeo de Derechos Humanos (CEDH).
El análisis exhaustivo del discurso de odio en las redes sociales y sus efectos difícilmente podría realizarse solo a través de una metodología doctrinal. Este tipo de análisis será particularmente beneficioso para comprender las implicaciones del discurso de odio que puede difundirse muy rápidamente sin restricciones en las fronteras nacionales y por cualquier persona, para comprender mejor los desafíos teóricos impuestos por el discurso de odio. Por lo tanto, se utilizará el método sociolegal para analizar los trabajos de comunicación, estudios sociales y de medios.
Se utiliza la amplia clasificación sugerida por el Comité de Ministros del Consejo de Europa (CoE), que considera "todas las formas de expresión que difunden, incitan, promueven o justifican el odio racial, la xenofobia, el antisemitismo u otras formas de odio basadas en la intolerancia (incluyendo: intolerancia expresada por nacionalismo agresivo y etnocentrismo, discriminación y hostilidad contra minorías, inmigrantes y personas de origen inmigrante)".
De las diferentes plataformas de redes sociales, Facebook y Twitter se encuentran entre las más populares con millones de usuarios. Las plataformas de colaboración en línea, como Google Docs, también se pueden considerar como una de red social que permiten que varias personas cooperen y transmitan un mensaje.

8. INFORMACIÓN GENERAL

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS

El objetivo de este discurso es ridiculizar a las víctimas, humillarlas y representar sus agravios como menos serios.

J.Waldron: 2 tipos de mensajes peligrosos. 1. Dirigido a las víctimas, pretende deshumanizarlas o ridiculizarlas y hacerlas sentir mal acogidas en la sociedad. El efecto global del delito de odio es insultar a las víctimas, estereotipándolas, por ejemplo, como terroristas, abogando por su exclusión de la sociedad, negándoles sus derechos humanos, responsabilizándolas por las acciones de otros miembros del grupo, aplicando estándares dobles etc. 2. Está dirigido al resto de la sociedad y tiene la intención de alentar a las personas de acuerdo con la idea de que ciertos grupos de la sociedad deberían ser excluidos y no tolerados.

Por otra parte, el discurso de odio una vez que aparece en línea, no desaparece, y abre la posibilidad de estar presente y reutilizado en las redes sociales por un período de tiempo ilimitado. Los estudios de los efectos del discurso de odio en línea muestran que el mayor peligro, sin embargo, puede provenir de la normalización del odio a través de las redes sociales. Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A. M., Bahfen, N., ... & Connelly, K. (2017). Ciber-racismo y resiliencia comunitaria: estrategias para combatir el odio racial en línea. Springer. p. 43 20 Ibid., p. 45.

Aparte de la controversia sobre el alcance de la aceptable libertad de expresión, es crucial entender que los derechos humanos y la libertad de expresion, en particular, fueron formulados en una época y contexto particulares cuando el factor de la redes sociales no existía.

La forma principal de odio que se practica comúnmente es la expresión directa de ideas negativas hacia los grupos objetivo. En ese caso, los usuarios se dirigen específicamente a ciertos grupos y difunden ideas que muestran hostilidad hacia ellos. Los grupos de odio no se limitan a las víctimas "tradicionales" del odio, como judíos, negros, musulmanes, homosexuales, sino que a veces se dirigen a víctimas más inusuales. Por ejemplo, el grupo de Facebook "Kick a Ginger Day" ha alentado los ataques físicos a los estudiantes con cabello rojo y, en consecuencia, tales ataques han ocurrido. Ambas páginas fueron eliminadas posteriormente por la administración de Facebook. Dichas plataformas brindan oportunidades ideales para odiar a los grupos, no simplemente difundir mensajes de odio, pero para transformar la comprensión del odio y hacer que sus mensajes sean más justificables. Basan el discurso de odio con supuestos hechos científicos. Ejemplos de tal odio podrían ser grupos antisemitas disfrazados de organizaciones de investigación de negación del Holocausto.

El Instituto de Prevención de Delitos de Odio ha reconocido otra forma de lavado de información relacionada con personas que fingen ser miembros de objetivos de discurso de odio. Este esquema funciona de la siguiente manera: páginas falsas pretenden promover, por ejemplo, agendas musulmanas, cuando en realidad, están publicando información que causaría una indignación pública. Comúnmente, tales páginas o cuentas falsas apoyan el terrorismo y la violencia. Un ejemplo, podría ser los eventos que tuvieron lugar (inmediatamente) después del asedio de Lindt Cafe en Sydney en 2014, dejando cuatro personas muertas. Después del ataque, varias páginas en Facebook fingieron ser musulmanes locales y expresaron su apoyo al ataque. Las páginas fueron retiradas después de que el Instituto de Prevención de Odio en línea informara tanto a Facebook como a la policía, sin embargo, el daño ya estaba hecho: la publicación fue vista por alrededor de 260 000 personas, dejando una falsa impresión sobre la comunidad musulmana.

9. INFORMACIÓN GENERAL

Nosotros y ellos: identificar el ciber-odio en Twitter mediante múltiples características protegidas. Pete Burnap y Matthew L Williams
https://doi.org/10.1140/epjds/s13688-016-0072-6© Burnap and Williams 2016
Recibido el 22 de noviembre de 2015; aceptado el 15 de marzo de 2016; publicado el 23 de marzo de 2016

10. INFORMACIÓN GENERAL

La extensión del discurso de odio en las redes sociales en línea.
Binny Mathew Ritam Dutt Pawan Goyal Animesh Mukherjee.
Instituto Indio de Tecnología, Kharagpur.
https://arxiv.org/abs/1812.01693

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

METODOLOGÍA
Se construyeron múltiples modelos individuales para clasificar el ciber odio en función de un rango de características protegidas tales como la raza, la discapacidad o la orientación sexual. Se usó un análisis del texto para extraer las dependencias escritas, que representan relaciones sintácticas y gramaticales entre palabras, posteriormente mostradas para capturar lenguaje 'exclusivo' – se mejoró consistentemente la clasificación de la máquina para los distintos tipos de ciber odio más allá del uso de una Bolsa de Palabras y los términos de odio conocidos, construidos por un modelo mixto basado en datos para mejorar la clasificación donde más de una característica protegida puede ser atacada (ej. La raza y la orientación sexual), contribuyendo al estudio incipiente de la interseccionalidad en los crímenes de odio.

11. INFORMACIÓN GENERAL

¿Pararse por Suecia? Los Discursos Racistas, Arquitecturas y Asequibilidades de un Grupo de Facebook Anti-Inmigración
Samuel Merrill Mathilda Åkerlund
Jornal de Comunicación Mediado por Ordenador, Volumen 23, Tema 6, 1 de noviembre de 2018, Páginas 332–353
https://doi.org/10.1093/jcmc/zmy018
Publicado: 27 septiembre de 2018 Artículo histórico
https://academic.oup.com/jcmc/article/23/6/332/5107230

12. INFORMACIÓN GENERAL

Inmigración, crisis económica y discursos radiofónicos: hacia un lenguaje excluyente
en Estudios sobre el Mensaje Periodístico vol. 20(2):899-916 · Enero 2014 con 29 lecturas.
DOI: 10.5209/rev_ESMP.2014.v20.n2.47063
https://www.researchgate.net/publication/271210973_Inmigracion_crisis_economica_y_discursos_radiofonicos_hacia_un_lenguaje_excluyente

13. INFORMACIÓN GENERAL

RED DE CENTROS DE ASESORAMIENTO PARA LAS VÍCTIMAS DEL RACISMO

La "Red de centros de asesoramiento para las víctimas del racismo" es un proyecto común de la asociación humanrights.ch y de la Comisión Federal contra el racismo CFR.

http://www.network-racism.ch/fr/accueil.html

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS

La red reúne actualmente a 24 servicios especializados de toda Suiza, que proporcionan consultas para los casos de discriminación racial. El objetivo principal de la Red de centros de asesoramiento para las víctimas del racismo es apoyar en su trabajo a los servicios asociados al proyecto. La Red de consultas publica el informe anual titulado "Incidentes racistas identificados por los centros de asesoramiento" y ofrece a sus miembros posibilidades de formación continua y actualización de redes.

METODOLOGÍA

Una vez al año, partiendo de la parte anonimizada del banco de datos DoSyRa, se elabora y publica un informe de evaluación sobre los incidentes señalados. Con este fin, los servicios de asesoramiento deben ordenar los casos, en el momento de recogida de los datos, en función de un marco analítico. Los informes de análisis se elaboran en base a esos datos. No se pretende una toma de datos exhaustiva. Se trata, más bien, de reflexionar y mostrar los expedientes que a diario se tratan en todos los centros de asesoramiento para las víctimas del racismo. Para registrar un caso en la base de datos, deben reunirse las siguientes condiciones: 1. Existencia de una interacción entre el centro de asesoramiento y la persona en cuestión; 2. Que la situación se haya descrito concretamente por el profesional competente considerándola como un caso de discriminación racial; 3. Que haya existido una consulta. El informe en cuestión presenta un análisis de los incidentes recogidos en 2017 y clasificados como casos de discriminación racial.

DESARROLLO Y ALCANCE DE LA INICIATIVA

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS

En el año 2017 se muestra claramente los casos de discriminación múltiple. Ser mujer, extranjera y negra, por ejemplo, puede considerarse especialmente grave. Frente a esto, lo relevante es mejorar la identificación de estos problemas, prevenirlos y corregirlos.

La escuela debería ser por excelencia el lugar donde los niños fueran protegidos ante todo tipo de discriminación. Sería absurdo pensar que eso es efectivamente así y el informe muestra que desgraciadamente no es el caso. Debemos, por tanto, preguntarnos cómo combatir mejor el fenómeno contando con que toda medida preventiva implica la motivación y la intervención de los profesionales en el ámbito escolar. Finalmente, la CFR publicó en 2017 un estudio y recomendaciones sobre el racismo anti-Negros. (**www.ekr.admin.ch/documentation/f107/1320.html**)

Como en años anteriores, la mayoría (192 de 301) de los incidentes identificados son comunicados por las propias víctimas. Áreas en las que la discriminación ha tenido lugar: el mundo del trabajo (43 incidentes) y área de la formación/escuela/ámbito familiar (42 incidentes), son los ámbitos donde más incide la discriminación. En la segunda categoría y a nivel de la escolarización obligatoria es donde se ha reseñado un nivel particularmente elevado de incidentes (31). Comparando con 2016, las discriminaciones han aumentado en 3 puntos porcentuales en las áreas de formación/escuela/ámbito familiar, relaciones de vecindad/barrio y servicios públicos proporcionados por los particulares. Formas de discriminación: en 2017 gran parte de los incidentes recogidos mostraban desigualdad en su tratamiento (107 casos=+7pp).

Prejuicios e ideologías como bases del incidente: sobre la xenofobia en general, la causa de discriminación más frecuentemente reseñada es el racismo antinegros (95 incidentes); la hostilidad respecto a las personas musulmanas es la tercera causa de discriminación con 54 casos (aumentando en 2 puntos porcentuales en relación con el año anterior). El número de incidentes contabilizados en la categoría similar relativa a racismo antiárabes también ha aumentado (+3pp). Discriminación múltiple: de 100 casos, los centros de asesoramiento han determinado que un tercio de ellos constituye una discriminación múltiple. Ésta se refiere principalmente a la nueva categoría de residente (28 incidentes).

14. INFORMACIÓN GENERAL

Van Dijk, T. Discurso Racista prólogo. P.9 -15 http://www.discursos.org/oldarticles/Discurso%20racista.pdf

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS

El discurso racista, junto con las otras prácticas (no verbales) discriminatorias, contribuye a la reproducción del racismo como una forma de dominación étnica o racial. Lo habitual es que se lleve a cabo mediante la expresión, confirmación o legitimación de las opiniones, actitudes e ideologías racistas del grupo étnico dominante. Aunque existen otros tipos de racismo en otras muchas partes del mundo, la forma de racismo más corriente e históricamente devastadora ha sido el racismo europeo contra los pueblos no europeos. Es por ello que este artículo se limitará al estudio del racismo europeo o "blanco", así como los diferentes tipos de discursos que giran a su alrededor. Discurso racista dirigido en contra de los otros. Básicamente, existen dos modalidades principales de discurso racista: a. Discurso racista dirigido a los Otros étnicamente diferentes. b. Discurso racista sobre los Otros étnicamente diferentes.

DESARROLLO Y ALCANCE DE LA INICIATIVA

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS

a. Discurso racista dirigido a los Otros étnicamente diferentes.

b. Considerado "políticamente incorrectos", la mayoría de los discursos racistas dirigidos a los miembros del grupo étnico dominado tienden a convertirse en sutiles e indirectos.

c. Discurso racista sobre los Otros étnicamente diferentes.

La característica general de este tipo de discurso racista se resume en una imagen negativa de Ellos, combinada frecuentemente con una representación positiva de nosotros. Algo típico es la negación o mitigación del racismo. Es posible que la inmigración sea tratada en términos de invasión, inundación, amenaza o, como un problema grave, en lugar de como una importante y necesaria contribución. El discurso de odio se basa en tres cuestiones principales: 1. Ellos son diferentes. 2. Ellos son perversos. 3. Ellos son una amenaza. Los Otros se representen en términos exóticos. Ellos son catalogados con idénticos patrones al pertenecer al mismo grupo (mientras que Nosotros somos todos diferentes individualmente unos de otros). Se destaca la perversidad del comportamiento de los Otros, que les lleva a romper y no cumplir nuestras normas y reglas. El tema más prominentemente tratado es la delincuencia.

Si estas personas de las minorías consiguen aparecer positivamente en las noticias, lo harán por haber destacado como campeones de algún deporte o como músicos. La lexicalización o selección de las palabras tiende a estar sesgada de muchas maneras, no sólo en el insulto racial o étnico explícito, sino también en formas más sutiles de discurso, empezando por el mismo problema de designar a los otros.

15. INFORMACIÓN GENERAL

Georgios K. Pitsilis, Heri RamampiaroHelge Langseth. 2018.
Detección efectiva del discurso de odio en datos de Twitter usando redes neurales recurrentes. Inteligencia aplicada, Volumen 48, Tema 12, pp 4730–4742
https://link.springer.com/article/10.1007/s10489-018-1242-y

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
El problema de entender el contenido de odio en las redes sociales. Se propone un esquema de detección como conjunto de clasificadores de Redes Neurales Recurrentes (RNN), incorporando varias características asociadas con información con relación al usuario, como su tendencia hacia el racismo o el sexismo.

METODOLOGÍA
Se evaluó el enfoque utilizado en un corpus de disponibilidad pública de 16000 tweets, y los resultados demostraron su efectividad en comparación con soluciones de la técnica existentes. Más específicamente, el esquema puede distinguir exitosamente entre mensajes de contenido racista y sexista, y los normales, así como conseguir una mayor calidad a la hora de clasificarlos en comparación con los algoritmos actualmente existentes.

16. INFORMACIÓN GENERAL

David RobinsonZiqi ZhangEmail autor Jonathan Tepper. 2018. Detección de Discurso de Odio en Twitter: Ingeniería de Características contra la Selección de Características. Conferencia Europea de Semántica en Web.
ESWC 2018: La Web Semántica: ESWC 2018 Eventos Satélite pp 46-49
https://link.springer.com/chapter/10.1007%2F978-3-319-98192-5_9

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

METODOLOGÍA
Se realizó un análisis de selección de características usando Twitter como caso de estudio, y se mostraron resultados que confrontan las percepciones tradicionales sobre la importancia de la ingeniería de selección de características manual con los modelos de selección automática que resultan mejores.

17. INFORMACIÓN GENERAL

Hajime Watanabe; Mondher Bouazizi; Tomoaki Ohtsuki. Discurso de odio en Twitter: Un Acercamiento Pragmático para Recoger Expresiones Odiosas y Ofensivas y Realizar una Detección del Discurso de Odio.
Acceso IEEE (Volumen: 6)
https://ieeexplore.ieee.org/document/8292838

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
El Discurso de Odio utiliza un lenguaje ofensivo, violento o agresivo, poniendo como objetivo a un grupo específico de personas que comparten una característica común, ya sea su género (ej. Sexismo), su grupo étnico o racial (ej. racismo), o sus creencias y religión. Mientras que la mayoría de las redes sociales y blogs prohíben el uso de discurso de odio, la cantidad de mensajes difundidos en estas redes y páginas web hacen casi imposible controlar todo su contenido. Es por eso por lo que aparece la necesidad de detectar automáticamente ese tipo de discurso para filtrar contenido que presente lenguaje odioso o que incite al odio. En este informe, se propone un acercamiento a la detección de expresiones de odio en Twitter.

17. INFORMACIÓN GENERAL

METODOLOGÍA

Basado en unigramas y patrones que son recogidos automáticamente del set de entrenamiento. Estos patrones y unigramas se usan posteriormente como, entre otras cosas, características para entrenar un algoritmo de aprendizaje. Los experimentos en un set de pruebas compuesto de 2010 tweets muestran que el acercamiento alcanza una precisión igual al 87,4% detectando si un tweet es ofensivo o no (clasificación binaria), y una precisión igual al 78,4% detectando si un tweet es odioso, ofensivo o limpio (clasificación terciaria).

Teniendo como base un conjunto de tweets, el objetivo de este trabajo es clasificar cada uno en una de las tres siguientes clases:

- Limpias: esta clase consiste en los tweets neutros, no ofensivos y que no presentan discurso de odio.
- Ofensivo: esta clase contiene tweets ofensivos, pero que no representan ningún odio o discurso racista/segregador.
- Odioso: esta clase incluye los tweets que son ofensivos, presentan odio, expresiones y palabras racistas y segregadoras.

Se usa el aprendizaje de la maquinaria para realizar la clasificación: se extrae un conjunto de características de cada tweet, se refieren a un set de entrenamiento y se realiza la clasificación. Se han recogido y combinado tres distintos tipos de datos:

- Un primer conjunto de datos disponible públicamente en Crowdflower2: este set de datos contiene más de 14000 tweets que han sido clasificados manualmente en una de las siguientes clases: "Odioso", "Ofensivo" y "Limpio". Todos los tweets de este set de datos han sido anotados manualmente por tres personas.
- Un segundo set de datos también disponible públicamente en Crowdflower3: siendo usado previamente en [19] y también anotado manualmente en una de las tres clases: "Odioso", "Ofensivo" y "Limpio".
- Un tercer set de datos, publicado en github4 y usado en el trabajo [18]: los tweets en este grupo de datos son clasificados en una de las tres clases siguientes: "Sexismo", "Racismo" y "Ninguno de ellos". Las dos primeras ("Sexismo" y "Racismo") se refieren a formas específicas de discurso de odio, incluidas como parte de la clase "Odioso", mientras que los tweets del grupo "Ninguno de ellos" han sido descartados porque no hay indicio de que sean o limpios u ofensivos (algunos tweets fueron revisados manualmente, identificados como ambas clases). Para realizar la tarea de clasificación, el grupo de datos se divide en tres subgrupos, los cuales son:
  - Un grupo de entrenamiento: este set contiene 21000 tweets, distribuidos de manera balanceada entre las tres clases (sean "Limpio", "Ofensivo" y "Odioso"): cada clase con 7000 tweets. Este set será referido como "grupo de entrenamiento" para el resto del trabajo.
  - Un grupo de prueba: este set contiene 2010 tweets, cada clase tiene 670 tweets. Este set será referido como "grupo de prueba" y se usará para optimizar nuestro acordado acercamiento.
  - Un grupo de validación: este set contiene 2010 tweets, cada clase con 670 tweets. Este set será referido como "grupo de validación" y será usado para evaluar nuestro acordado acercamiento.

17. INFORMACIÓN GENERAL

1. Características en base al sentimiento. Para ello, se extraen las siguientes características de cada tweet:
- El puntaje total de palabras positivas (PW),
- El puntaje total de palabras negativas (NW),
- El rango de palabras con relación a un sentimiento (positivo o negativo) p(t) definido como: $\rho(t) = \frac{PW-NW}{PW+NW}$; P(t) iguala a 0 si el tweet no tiene palabras que expresen un sentimiento,
- El número de palabras de una jerga positiva,
- El número de palabras de una jerga negativa,
- El número de emoticonos positivos,
- El número de emoticonos negativos,
- El número de hashtags positivos,
- El número de hashtags negativos.

2. Características semánticas
- El número de signos de exclamación,
- El número de signos de interrogación,
- El número de puntos y final,
- El número de palabras en mayúscula,
- El número de citas,
- El número de interjecciones,
- El número de expresiones cómicas,
- El número de palabras en un tweet.

3. Características de unigrama

Las características de unigrama son unigramas simples recogidos por el grupo de entrenamiento de una forma pragmática, usados cada uno como una característica independiente que puede tener dos valores: "verdadero" y "falso".

Mientras que la mayoría de las palabras correspondientes a ambas clases son palabras generales que las personas utilizan para insultar o degradar a alguien, algunas de ellas tienen contenido racista o que se refiere a una característica específica de género, grupo étnico, u otros (ej. 'musulmanes', 'islámico', 'maricón', 'sudaca', etc.)

4. Características de diseño.

DESARROLLO Y ALCANCE DE LA INICIATIVA

17. INFORMACIÓN GENERAL

PRINCIPALES DISCURSOS DE ODIO Y FACTORES IDENTIFICADOS

En el contexto de internet y las redes sociales, el discurso de odio no solo crea tensión entre grupos de población, sino también su impacto puede afectar a los negocios, o iniciar serios conflictos en la vida real. Es por ello por lo que, normas de estilo de Facebook, Youtube y Twitter prohíben el uso de discurso de odio. Sin embargo, siempre es difícil controlar y filtrar todo el contenido. Por lo tanto, viéndolo desde un enfoque de estudio e investigación, el discurso de odio siempre se ha intentado detectarlo automáticamente. La mayoría de esas labores de detección tienen como objetivos la creación de diccionarios de palabras de odio y expresiones [4] o una clasificación binaria entre "odio" y "no odio" [5]. Sin embargo, es difícil discernir de una manera clara si una frase tiene odio o no, sobre todo si el discurso de odio esconde sarcasmo o si no hay signos claros que muestren odio, racismo o estereotipos existentes.

Se puede debatir que la técnica de análisis basada en el sentimiento sea válida para detectar discurso de odio. Sin embargo, hablamos de una tarea distinta, que requiere técnicas más sofisticadas: en el análisis sentimental, el principal objetivo es detectar polaridad de sentimientos en los tweets, tiene conexión con la idea de detectar cualquier palabra o frase de ámbito positivo o negativo.

En un contexto relacionado, los patrones de escritura han demostrado ser efectivos para tareas de clasificación de texto del tipo detección de sarcasmo [6], [7], análisis de distinta clase de sentimientos [8], o cuantificación de sentimientos [9]. El tipo de patrones, y la manera en la que están elaborados, se construyen y extraen dependiendo de la aplicación.

1. Se propone un acercamiento a los patrones para detectar discurso de odio en Twitter: los patrones se extraen de una manera pragmática del grupo de entrenamiento y se define un conjunto de parámetros para optimizar la recolecta de patrones.
2. En adición a los patrones, se propone recolectar de una manera pragmática también palabras y expresiones que muestren odio y ofensa, para usarlas junto con los patrones y otras características basadas en el sentimiento para detectar discurso de odio.
3. El grupo propuesto de unigramas y patrones se podrán usar como diccionarios para futuros trabajos de detección de discurso de odio.

Se clasifican tweets en tres clases distintas (en vez de dos), haciendo distinción entre los tweets que muestran odio, y aquellos que simplemente pretenden ofender.

Para concluir, principalmente son 4 tipos de características las que se pueden extraer: "características en base a los sentimientos", "características semánticas", "características del unigrama" y "características de diseño o patrón". Combinando estos sets mencionados, es posible que se detecte el discurso de odio: "las características en base a los sentimientos" permiten extraer la polaridad de un tweet, un componente esencial del discurso de odio (debido a que la mayoría de los discursos del odio son negativos). "Las características semánticas" permiten encontrar una expresión enfatizada. "Las características del unigrama" permiten detectar cualquier forma explícita de discurso de odio, mientras que los patrones o diseños permiten identificar cualquier otra forma, ya sea implícita o no, de discurso de odio.

IMPACTO (ámbito de influencia de la iniciativa)

Los resultados de la investigación han sido posibles gracias al trabajo de "Seguridad Cognitiva: Un Nuevo Acercamiento para Asegurar a Grande Escala la Distribución de Aplicaciones para el Móvil", la Comisión de Investigación del Instituto Nacional de Tecnologías de la Información y la Comunicación (NICT), JAPÓN.

18. INFORMACIÓN GENERAL

Lev-On, A. 2018. ¿La red anti-social? Encuadrando las redes sociales en tiempos de Guerra. Redes Sociales + Sociedad, 4(3).
https://journals.sagepub.com/doi/full/10.1177/2056305118800311

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Análisis de contenido de nuevos artículos que incluyen referencias a las redes sociales en seis periódicos diarios en Israel escritos en hebreo durante la Guerra Israel-Gaza (2014). Los informes enmarcaron las redes sociales como espacios para el discurso de odio y la distribución de rumores. Adicionalmente se analizaron las redes sociales como canales de comunicación alternativos para los políticos y las celebridades, como zonas de uso para hacer diplomacia pública. Las redes sociales raramente se consideraban como plataformas para orquestar acciones colectivas o enfrentarse al enemigo.

19. INFORMACIÓN GENERAL

Perello Camacho, Carlos. Detección multilenguaje del Discurso de Odio contra las mujeres e inmigrantes en Twitter.
Licenciado en Ingeniería Computacional.
https://rua.ua.es/dspace/bitstream/10045/93563/1/Multilingual_Detection_of_Hate_Speech_Against_Immigra_Perello_Camacho_Carlos.pdf

CARACTERÍSTICAS GENERALES DE LA INICIATIVA

OBJETIVOS
Debido al masivo incremento de usuarios de las redes sociales, la presencia de abuso verbal, discursos de odio y actitudes de bullying han crecido también en estos años. Especialmente en Twitter, donde los usuarios encuentran una manera de acosar y ofender a otros individuos y colectivos anónimamente, y no hay suficientes métodos para impedirlo. Este proyecto describe la implementación de un sistema de detección de discurso de odio contra las mujeres y los inmigrantes con el objetivo de reducir el odio en las redes sociales, así como participar en el SemEval-2019 Task 5 challenge. SemEval-2019 Task 5 consiste en detectar el discurso de odio contra las mujeres y los inmigrantes en Twitter, tanto en inglés como en español. Para ello, se usaron técnicas tradicionales de Aprendizaje de Maquinaria *(Machine Learning* en inglés). Nuestro sistema se basa principalmente en el uso de n-gramas, análisis de opiniones y palabras.
Nuestro sistema obtuvo la segunda mayor precisión en la Tarea A en español, sobrepasando los sistemas más complejos de redes neutrales entre un total de 40 participantes.

7.2. Webgrafía

http://plataformaciudadanacontralaislamofobia.org/
http://shr.gs/SJeLPYC
http://www.selfdefenceit.maiz.at/index9edd.html?q=es/content/self-defence-it
https://m.eldiario.es/redaccion/Desalambre-Malditaes-colaboraran-desmentir-inmigracion_6_857124301.html
https://www.opensocietyfoundations.org/voices/hate-crimes-don-t-discriminate
http://www.facingfacts.eu/
https://www.react-to-racism.brussels/
https://www.opensocietyfoundations.org/explainers/open-society-foundations-spain
https://fundraising.co.uk/2016/02/02/donate-the-hate-turns-hates-comments-into-donations/#.XEhTXXAzUpE
https://blogextranjeriaprogestion.org/2018/12/24/guia-practica-sobre-delitos-de-odio/
https://saveahater.accem.es/
https://www.accem.es/wp-content/uploads/2018/12/impacto-de-las-brechas-digitales-en-la-poblacion-extranjera.pdf
http://www.alertadiscriminacion.org/es/pagina-principal
http://www.equineteurope.org/Not-on-Our-Watch-Equality-Bodies-fighting-Hate-Speech
https://www.ihrec.ie/documents/hatetrack-tracking-and-monitoring-racist-hate-speech-online/
http://elpais.com/internacional/2016/12/05/mundo_global/1480955519_917792.html
https://www.lavanguardia.com/local/lleida/20181109/452790530985/campana-jovenes-musulmanas-islamofobia.html
http://asceps.org/words-are-stones/
https://www.cear-euskadi.org/19-herramientas-contra-el-discurso-del-odio/
http://www.somos-mas.es/?utm_source=Newsletter%20J%C3%B3venes%20y%20Desarrollo&utm_campaign=ec085fc95a-EMAIL_CAMPAIGN_2018_02_13&utm_medium=email&utm_term=0_e3c739c300-ec085fc95a-80311961
http://hatemeter.eu/
http://www.voxdiritti.it/ecco-le-mappe-di-vox-contro-lintolleranza/,
http://www.nohatespeech.it/
https://www.silencehate.it/
http://hatespeech.di.unito.it/
https://www.amnesty.it/entra-in-azione/discorsi-dodio-online-combattiamoli-insieme/
(https://www.senato.it/service/PDF/PDFServer/DF/338344.pdf)
https://www.imra.org.il/story.php3?id=73065
https://www.researchgate.net/publication/328966730_Hate_is_in_the_air_But_where_Introducing_an_algorithm_to_detect_hate_speech_in_digital_microenvironments
http://www.dtbg.nl/org/En/index.html
http://humanrightsutrecht.nl/ (also in English)
https://plan-einstein.nl/
http://www.inclusiveworks.eu/Portals/0/Hate%20Speech/hate%20speech%20downloadable%20versie_paginas_def2.pdf
http://www.miramedia.nl/projecten/digitaal-burgerschap.htm

GOBIERNO DE ESPAÑA
MINISTERIO DE TRABAJO, MIGRACIONES Y SEGURIDAD SOCIAL
SECRETARÍA DE ESTADO DE MIGRACIONES
SECRETARÍA GENERAL DE INMIGRACIÓN Y EMIGRACIÓN
DIRECCIÓN GENERAL DE INTEGRACIÓN Y ATENCIÓN HUMANITARIA
oberaxe Observatorio Español del Racismo y la Xenofobia

Cofinanciado por el Programa Derechos, Igualdad y Ciudadanía de la Unión Europea

ÍNDICE