PROTOCOLO Y SISTEMA DE INDICADORES

para la detección del discurso de odio en las redes sociales

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 INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES

oberaxe
Observatorio Español del Racismo y la Xenofobia

Organizaciones del Consorcio AL-RE-CO y autores/as:

CREA, Community of Researchers on Excellence for All
Lena de Botton, Joan Cabré, Cristina Pulido.

Con la colaboración de:

Maria Padrós, Teresa Plaja, Ana Toledo.
OBERAXE – MITRAMISS
MINISTERIO DEL INTERIOR
ASOCIACION TRABE

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: http://www.inclusion.gob.es/oberaxe/es/index.htm

NIPO PDF: 121-20-007-5
Diseño y maquetación: Carmen de Hijes

Esta publicación ha sido producida con el apoyo financiero del Programa Derechos, Igualdad y Ciudadanía de la Unión Europea. Los contenidos de esta publicación son responsabilidad de los socios del proyecto ALRECO y no reflejan las opiniones de la Comisión Europea.

ÍNDICE

Marco conceptual y enfoque de género

Metodología

Banco de palabras e indicadores

Anexos

MARCO CONCEPTUAL Y ENFOQUE DE GÉNERO

1

Introducción

1. Introducción

El proyecto AL-RE-CO (Discurso de odio, racismo y xenofobia: mecanismos de alerta y respuesta coordinada) **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 el racismo, la xenofobia, la islamofobia y el antisemitismo. El proyecto esta co-financiado por la Dirección General de Justicia de la Comisión Europea, en el marco del Programa de Derechos, Igualdad y Ciudadanía de la Unión Europea y es desarrollado por un consorcio de cuatro instituciones yentidades, coordinado por el Observatorio Español del Racismo y la Xenofobia (OBERAXE), de la Dirección General de Integración y Atención Humanitaria, de la Secretaría de Estado de Migraciones, en el Ministerio de Trabajo, Migraciones y Seguridad Social.**

El proyecto lleva a cabo tres grandes líneas de trabajo:

1. Desarrollar un Protocolo y sistema de indicadores, que incluye un Informe de Buenas Prácticas
2. Desarrollo de una herramienta informática para la detección y seguimiento del discurso de Odio (en Twitter)
3. Diseño de Estrategias compartidas entre actores, instituciones y agentes clave, incluidos los colectivos afectados por el discurso de odio.

En una primera fase de trabajo, se ha realizado una identificación de experiencias y buenas prácticas, desarrolladas principalmente en la Unión Europea, que permiten avanzar en el desarrollo de un protocolo y del sistema de indicadores. Fruto de esta revisión de experiencias a nivel internacional (proyectos, informes, plataformas, herramientas y literatura científica) se elaboró el **"Informe de buenas prácticas o experiencias similares desarrolladas en la UE para identificar, analizar, monitorizar y evaluar el discurso de odio en línea, por motivos racistas, xenófobos, islamófo- bos y antisemitas"**, que constituye el primer producto del proyecto (D 2.1).

Como segundo producto, se presenta a continuación el "**Protocolo y Sistema de indicadores para la detección del discurso de odio en línea"**. Su objetivo es desarrollar un protocolo de actuación que contenga un sistema de indica- dores, con criterios de búsqueda, sobre discursos que fomenten el racismo, la xenofobia y el odio en la red. El sistema incluirá también **indicadores de alerta temprana** que permitan evaluar la intensidad, gravedad, distribución, y poten- cial impacto del discurso de odio, con el fin de establecer recomendaciones de acción para prevenir posibles incidentes discriminatorios o delitos de odio.

2

Marco conceptual: el discurso de odio

2. Marco conceptual: el discurso de odio

El discurso de odio se ha reforzado y expandido por medio de Internet, y ello ha conducido de algún modo a una normalización del mismo. Esta difusión en línea del discurso de odio esté permitiendo que llegue fácilmente al conjunto de la ciudadanía, especialmente a las personas jóvenes. Las organizaciones, grupos o personas que difunden ideas racistas, xenófobas, islamófobas y antisemitas, han encontrado de este modo nuevas vías para hacerlo y su impacto es cada vez mayor. Puede afirmarse que la difusión en línea del discurso de odio ha supuesto un salto cualitativo en todos los países europeos, especialmente a través de las redes sociales.

Según el 4º Informe de Monitoreo de la Comisión Europea sobre el Código de Conducta para la lucha contra el discurso de odio ilegal en la Red[1] -enero 2019-, el discurso de odio racista y xenófobo, principalmente dirigido contra las personas inmigrantes, refugiadas y minorías étnicas, continúa siendo predominante y el más denunciado en las redes sociales. En este sentido, de las denuncias presentadas en el año 2018, el 17% correspondían a discursos de odio xenófobos (anti-inmigrantes), el 15,6% por orientación sexual, el 13% por islamofobia, el 12,2% por anti-gitanismo y el 10,1% correspondía a discursos antisemitas.

A esta situación, se une el amplio debate existente entre las definiciones jurídicas, conceptuales, ideológicas y de impacto social del discurso de odio. Se habla de "discurso de odio", "delito de odio", "crimen de odio", "incidente de odio", etc. Igualmente, los términos "racismo", "xenofobia", "islamofobia" y "antisemitismo" no están exentos de discusión e interpretaciones. A efectos de este proyecto, por ello, se hace necesario establecer un **marco conceptual**, previo al desarrollo del Protocolo y los Indicadores, que se incluyen en este documento, nos basaremos en un conjunto de definiciones establecidas y consensuadas en la legislación, normativa y compromisos nacionales e internacionales, y por otro lado, en el marco del Acuerdo suscrito entre el Consejo General del Poder Judicial, la Fiscalía General del Estado, el Ministerio de Justicia, el Ministerio de Interior, el Ministerio de Educación y Formación Profesional, el Ministerio de Trabajo, Migraciones y Seguridad Social, el Ministerio de la Presidencia, Relaciones con las Cortes e Igualdad, el Ministerio de Cultura y Deporte y el Centro de Estudios Jurídicos para Cooperar Institucionalmente en la lucha contra

1 Accesible en https://ec.europa.eu/info/sites/info/files/code_of_conduct_factsheet_7_web.pdf

el Racismo, la Xenofobia, la LGBTIfobia y otras formas de intolerancia[2]. En primer lugar, la Recomendación nº R (97) 20, de 30 de octubre de 1997, del Comité de Ministros del Consejo de Europa sobre el discurso de odio, en la que define que: "se entenderá por discurso de odio todas las formas de expresión que difundan, inciten, promuevan o justifiquen el odio racial, la xenofobia, el antisemitismo u otras formas de odio basadas en la intolerancia, incluyendo la intolerancia expresada por el nacionalismo agresivo y el etnocentrismo, la discriminación y la hostilidad contra las minorías, los migrantes y las personas de origen inmigrante".

La Recomendación de Política General nº 15 relativa a la lucha contra el discurso de odio y su Memorándum explicativo, de la Comisión Europea contra el Racismo y la Intolerancia (ECRI), del Consejo de Europa[3]. Según la ECRI, **el discurso de odio** debe entenderse como "fomento, promoción o instigación, en cualquiera de sus formas, del odio, la humillación o el menosprecio de una persona o grupo de personas, así como el acoso, descrédito, difusión de estereotipos negativos, estigmatización o amenaza con respecto a dicha persona o grupo de personas y la justificación de esas manifestaciones por razones de raza, color, ascendencia, origen nacional o étnico, edad, discapacidad, lengua, religión o creencias, sexo, género, identidad de género, orientación sexual y otras características o condiciones personales". Dado que todos los seres humanos pertenecen a la misma especie, la ECRI rechaza las teorías que sostienen la existencia de distintas razas. Sin embargo, en esta Recomendación, la ECRI emplea el término "raza" a fin de garantizar que las personas que suelen percibirse de forma general y errónea como pertenecientes a otra raza queden sujetas a la protección que confiere dicho texto.

2 Acuerdo suscrito entre el Consejo General del Poder Judicial, la Fiscalía General del Estado, el Ministerio de Justicia, el Ministerio del Interior, el Ministerio de Educación y Formación Profesional, el Ministerio de Trabajo, Migraciones y Seguridad Social, el Ministerio de la Presidencia, Relaciones con las Cortes e Igualdad, el Ministerio de Cultura y Deporte y el Centro de Estudios Jurídicos para cooperar institucionalmente en la lucha contra el racismo, la xenofobia, la LGTBifobia y otras formas de intolerancia. Accesible http://www.inclusion.gob.es/oberaxe/ficheros/ejes/cooperacion/Acuerdo_insterinsticuional_original.pdf

3 ECRI RPG n15 relativa a la lucha contra el discurso de odio y su Memorandum explicativo. Accesible en http://www.inclusion.gob.es/oberaxe/ficheros/documentos/2016_12_21-Recomendacion_ECRI_NO_15_Discurso_odio-ES.pdf

A su vez, la **OSCE** en su Consejo Ministerial (diciembre 2003) invitó a los Estados a incorporar de una u otra forma los delitos de odio y a elaborar información fidedigna y estadísticas (Decisión sobre Tolerancia y No Discriminación nº4/3). Acordó el concepto **Delito de Odio** para hacer referencia al **delito motivado por intolerancia**, es decir, por **prejuicio o animadversión** hacia la víctima. OSCE (2003) definiéndolo como: "toda infracción penal, incluidas las infracciones contra las personas y la propiedad, cuando la víctima, el lugar o el objeto de la infracción son seleccionados a causa de su conexión, relación, afiliación, apoyo o pertenencia real o supuesta a un grupo que pueda estar basado en la "raza", origen nacional o étnico, el idioma, el color, la religión, la edad, la minusvalía física o mental, la orientación sexual u otros factores similares, ya sean reales o supuestos".

Como ya se ha señalado, el proyecto AL-RE-CO se centra en el discurso motivado por racismo, xenofobia, islamofobia y antisemitismo. No obstante, a efectos prácticos se ha considerado en el marco del racismo, la referencia al concepto de antigitanismo[4] señalado específicamente. A continuación, se recogen las siguientes conceptualizaciones:

4 Accesible en **https://rm.coe.int/ecri-general-policy-recommendation-no-13-on-combating-anti-gypsyism-an/16808b5aef**

Racismo

Se refiere a la creencia de que, por motivo de la raza, el color, el idioma, la religión, la nacionalidad, el origen nacional o étnico, se justifica el desprecio de una persona o grupo de personas o la noción de superioridad de una persona o grupo de personas. *(RPG No 7. Aunque la religión no aparece en la definición de discriminación racial en el artículo 1 de la* ***Convención Internacional sobre la Eliminación de todas las formas de Discriminación Racial***[5]*, el Comité para la Eliminación de la Discriminación Racial, reconoce, a la luz del principio de interseccionalidad que, el discurso de odio se extiende al discurso " dirigido contra las personas pertenecientes a determinados grupos étnicos que profesan o practican una religión distinta de la mayoría, por ejemplo las expresiones de islamofobia, antisemitismo y otras manifestaciones de odio similares contra grupos etnorreligiosos, así como las manifestaciones extremas de odio tales como la incitación al genocidio y al terrorismo". Recomendación General No 35 sobre la Lucha contra el discurso de Odio Racista, CERD/C/GC/35, 26 de septiembre de 2013, apartado 6.*

Antigitanismo[6]

Se refiere a una forma específica de racismo contra la población Roma/Gitanos-as y se define como "la ideología basada en la superioridad racial, una forma de deshumanización y de racismo institucional alimentado por una discriminación histórica, que se manifiesta, entre otras cosas, por la violencia, el discurso del miedo, la explotación y la discriminación en su forma más flagrante"[7].

5 Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, adoptada y abierta a la firma y ratificación por la Asamblea General en su resolución 2106 A (XX), de 21 de diciembre de 1965, en vigor desde el 21 de diciembre de 1969. Accesible en **https://www.ohchr.org/SP/ProfessionalInterest/Pages/CERD.aspx**

6 ECRI.RPG nº 13 sobre lucha contra el antigitanismo y las discriminaciones contra los romaníes/gitanos adoptada el 24 de junio de 2011 p 4.

7 "El Grupo de Alto Nivel de la Comisión Europea para combatir el Racismo y la Xenofobia ha añadido como categoría el antigitanismo" en el seguimiento del discurso de odio en la UE **https://www.gitanos.org/actualidad/archivo/125684.html.es**

Xenofobia

Se refiere al prejuicio contra, el odio hacia o el miedo a personas de otros países o culturas. Sentimiento, actitud o comportamiento de hostilidad, rechazo u odio hacia personas extranjeras o percibidas como tales. Es un prejuicio etnocentrista hacia la cultura, valores y tradiciones del extranjero, y se manifiesta desde el rechazo más o menos obvio, el desprecio y las amenazas, segregación, privación de derechos, hasta las agresiones y asesinatos. En el reciente Pacto Mundial para la Migración Segura, Ordenada y Regular de la ONU[8], de diciembre de 2018, en el marco de cooperación, en su punto número 15 dispone que el Pacto Mundial se basa en las personas, con una importante dimensión humana, y en su apartado f) señala de forma literal que "El Pacto Mundial se basa en el derecho internacional de los derechos humanos y defiende los principios de no regresión y no discriminación. La aplicación del Pacto Mundial asegurará el respeto, la protección y el cumplimiento efectivos de los derechos humanos de todos los migrantes, independientemente de su estatus migratorio, durante todas las etapas del ciclo de la migración. También reafirmamos el compromiso de eliminar todas las formas de discriminación contra los migrantes y sus familias, como el racismo, la xenofobia y la intolerancia". Además, el punto número 16 de este Pacto Mundial establece que en la Declaración de Nueva York para los Refugiados y Migrantes se aprobó una declaración política y un conjunto de compromisos, que sirvió para establecer el marco de cooperación sobre el que se asienta el Pacto Mundial, que consta de 23 objetivos con sus medidas de aplicación. Entre esos objetivos, en el número 17 se establece: "Eliminar todas las formas de discriminación y promover un discurso público con base empírica para modificar las percepciones de la migración".

8 Conferencia Intergubernamental encargada de aprobar el Pacto Mundial para la Migración Segura, ordenada y Regular. Marrakech (Marruecos), 10 y 11 de diciembre de 2018. Documento final de la conferencia. A/CONF.231/3 Accesible en **https://undocs.org/es/A/CONF.231/3**

Antisemitismo

El 26 de mayo de 2016, 31 países miembro de la Alianza Internacional para el Recuerdo del Holocausto, IHRA9, por sus siglas en inglés, adoptaron la definición de antisemitismo como sigue:

"El antisemitismo es una cierta percepción de los judíos que puede expresarse como el odio a los judíos. Las manifestaciones físicas y retóricas del antisemitismo se dirigen a las personas judías o no judías y/o a sus bienes, a las instituciones de las comunidades judías y a sus lugares de culto".[10]

La UE "pide a los Estados miembros que todavía no lo hayan hecho que refrenden la definición operativa de "antisemitismo" no vinculante jurídicamente que emplea la Alianza Internacional para el Recuerdo del Holocausto (IHRA) como una herramienta útil de orientación en la educación y la formación, también para las autoridades encargadas de hacer cumplir la ley a la hora de identificar e investigar los ataques antisemitas con mayor eficiencia y eficacia, y en el mismo sentido, se recoge en la Resolución del Parlamento Europeo sobre la lucha contra el antisemitismo. [11]

Por su parte, el Consejo Permanente Nº 1214 de la OSCE[12] el mes de febrero de 2019, con motivo del Día Internacional de Conmemoración del Holocausto se pronunció en idéntico sentido.

Igualmente, en el informe de la ONU A/74/358 de 23 de septiembre de 2019 sobre la "Eliminación de todas las formas de intolerancia religiosa" el Relator Especial sobre la libertad de religión o de creencias identifica la violencia, la discriminación y las expresiones de hostilidad incluidas las realizadas en línea, y motivadas por antisemitismo como un obstáculo relevante para el ejercicio del derecho a la libertad de religión y creencias, al tiempo que insta de forma urgente a los Estados que adopten un enfoque de derechos humanos para combatir el antisemitismo, y otras formas de intolerancia religiosa. En este documento se aborda el antisemitismo y su caracterización desde la perspectiva religiosa.[13]

9 Reúne a Gobiernos y expertos a fin de reforzar, impulsar y promover la educación, la memoria y la investigación en todo el mundo sobre el Holocausto, así como de mantener los compromisos de la Declaración de Estocolmo de 2000.

10 En los siguientes enlaces se encuentran ejemplos contemporáneos descritos en la vida pública, en los medios de comunicación, en las escuelas, en el lugar de trabajo y en la esfera religiosa: https://www.holocaustremembrance.com/es/node/196 y https://ep-wgas.eu/docs/pdf/IHRA-Definition-of-Antisemitism-ES-web.pdf

11 Resolución PE de 1 de junio de 2017 sobre lucha contra el antisemitismo (2017/2692 (RSP)) DOUE 30.8.2018, serie C 307/183, p. 184. Accesible en https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC_2018_307_R_0030&from=ES

12 OSCE Conseil Permanent Nº 1214 Vienne, Déclaration de l'UE à l'occasion de la Journée internationale dédiée à la mémoire des victimes de l´Holocauste, 31 janvier 2019 p.3 Accesible en https://www.osce.org/fr/permanent-council/412151?download=true

13 ONU A/74/358, 23 de septiembre de 2019 'Elimination of all forms of religious intolerance' Accesible en https://www.ohchr.org/Documents/Issues/Religion/A_74_47921ADV.pdf

Islamofobia

Desde el punto de vista etimológico y conceptual se refiere a la aversión hacia el islam, los musulmanes o lo musulmán.

Para la OSCE y la UNESCO[14] en el documento "Directrices para educadores sobre la manera de combatir la intolerancia y la discriminación contra los musulmanes, afrontar la islamofobia mediante la educación" utilizan la definición general "intolerancia y discriminación contra los musulmanes", ya que es el más utilizado por las organizaciones intergubernamentales, entre ellas, la propia OSCE, la UNESCO y el Consejo de Europa. Hay otros términos que hacen referencia a la intolerancia y discriminación hacia los musulmanes, como "islamofobia" o "racismo antimusulmán". La palabra "islamofobia" muy utilizada por las ONG y frecuentemente en los medios de comunicación, se entiende como temor, odio o prejuicio respecto del islam y los musulmanes. El concepto "racismo antimusulmán" sitúa la intolerancia frente a los musulmanes en el marco más amplio del racismo, e implica una interpretación racial de un concepto religioso. El término subraya el aspecto pluridimensional de la intolerancia frente a los musulmanes, que puede estar basado en factores distintos de la religión. Aunque estos términos no son sinónimos y hacen referencia a diferentes aspectos del problema, se utilizan con frecuencia indistintamente".

El Consejo de Europa[15] define islamofobia como "(…) el temor o los prejuicios hacia el islam, los musulmanes y todo lo relacionado con ellos. Tome la forma de manifestaciones cotidianas de racismo y discriminación u otras formas más violentas, la islamofobia constituye una violación de derechos humanos y una amenaza para la cohesión social".

Por otro lado, la islamofobia se ha definido como el sentimiento y actitud de rechazo y hostilidad hacia el islam y, por extensión, a las personas musulmanas. Reconocido por la FRA, la Agencia Europea de Derechos Fundamentales, a partir de la organización británica Runnymede Trust que elaboró el contenido del concepto, reconoció que hay ocho características que denotan islamofobia: la creencia de que el islam es un bloque monolítico, estático y refractario al cambio, radicalmente distinto de otras religiones y culturas con las que no comparte valores o influencias; inferior a la "cultura occidental" (primitivo, irracional, bárbaro y machista); violento y hostil per se; la ideología política y la religión están íntimamente unidas; el rechazo global a las críticas a Occidente formuladas desde ámbitos musulmanes; y la consideración de dicha hostilidad como algo natural y habitual.

14 OSCE UNESCO "Directrices para educadores sobre la manera de combatir la intolerancia y la discriminación contra los musulmanes: afrontar la islamofobia mediante la educación". Accesible en **https://www.osce.org/es/odihr/91301?download=true**

15 Consejo de Europa, 'Islamophobia and its consequences on Young People, European Youth Centre'. Budapest, 1-6 june 2004, informe a cargo de Ingrid Ramberg. Accesible en **http://www.observatorioislamofobia.org/que-es-la-islamofobia/**

Por su parte, el "Informe de Delimitación Conceptual en materia de delitos de odio"[16] elaborado en el marco de la Comisión de Seguimiento del citado Acuerdo suscrito entre el Consejo General del Poder Judicial, la Fiscalía General del Estado, el Ministerio de Justicia, el Ministerio de Interior, el Ministerio de Educación y Formación Profesional, el Ministerio de Trabajo, Migraciones y Seguridad Social, el Ministerio de la Presidencia, Relaciones con las Cortes e Igualdad, el Ministerio de Cultura y Deporte y el Centro de Estudios Jurídicos para Cooperar Institucionalmente en la lucha contra el Racismo, la Xenofobia, la LGBTIfobia y otras formas de intolerancia, señalaba que: "e**l delito de discurso de odio** (hate speech crime), respetando la definición de la Recomendación R (97) 20 del Comité de Ministros del Consejo de Europa y la posterior de la ECRI, es todo aquel acto de habla (i.e., manifestación **expre- sivo-comunicativa**) sancionado penalmente que pueda considerarse delito de odio. Los delitos de discurso de odio consisten paradigmáticamente en delitos de peligro o de clima (que favorecen, por ejemplo, la comisión de otros delitos de odio que no consistan en discurso), en los que se sanciona la incitación al odio y a la violencia, como por ejemplo la conducta tipificada por el artículo 510 del Código Penal (CP) español. Nada obsta, sin embargo, a que puedan considerarse delitos de discurso de odio otras conductas típicas consistentes en actos de habla (expresivo-comunicativo), como un delito de injurias agravado por la vía del artículo 22.4ª CP".

El discurso de odio (hate speech) se refiere a actos de habla con un contenido expresivo-comunicativo de odio o prejuicio del autor hacia determinada persona por razón de una condición personal, o que generan un efecto discriminatorio en un colectivo caracterizado por una condición personal. Por tanto, el discurso de odio puede constituir un delito penal o no. Un discurso de odio no requiere estar sancionado penalmente para poder denominarse de esta manera; sin embargo su impacto social es siempre negativo puesto que está directamente relacionado con los prejuicios, estereotipos y discriminación hacia determinados grupos o personas.

Los incidentes de odio conforme a la orientación de la RPG nº11 de la ECRI, es cualquier incidente que se percibe como delito de odio por la víctima o cualquier otra persona. Son hechos que, pudiendo ser indiciariamente constitutivos de un delito de odio o de una infracción administrativa relacionada con un delito de odio, no son delito: ya sea porque no son constitutivos de infracción alguna, ya sea porque sólo son constitutivos de infracción administrativa, ya sea porque todavía no se ha dictado sentencia condenatoria por la comisión del delito de odio en cuestión. Un discurso de odio que no pueda catalogarse como delito de discurso de odio sí puede ser, por tanto, tildado de incidente de odio.

Para AL-RE-CO, cuyo objetivo es dotarse de herramientas que faciliten la identificación y monitorización del discurso de odio racista, xenófobo, islamófobo, antisemita y antigitano, es importante también establecer algunas consideracio-

16 DÍAZ LÓPEZ, JA. "Informe de Delimitación Conceptual en materia de delitos de odio", marzo 2018. Ed. Ministerio de Empleo y Seguridad Social. Accesible en http://www.inclusion.gob.es/oberaxe/ficheros/documentos/InformeConceptualDelitosOdio.pdf

nes sobre los **Indicadores** que pueden aplicarse para identificar el discurso de odio, así como los criterios y factores que pueden disparar las alertas sociales, de cara a prevenir o contrarrestar incidentes y delitos de odio.

Según la Recomendación Nº 15 de la ECRI, un rasgo característico del uso del discurso de odio es que puede tener la intención de incitar a otros a cometer actos de violencia, intimidación, hostilidad o discriminación contra aquellos a quienes va dirigido, o cabe razonablemente esperar que tenga tal efecto. E**l elemento de incitación** significa que, o bien existe una intención clara de cometer actos de violencia, intimidación, hostilidad o discriminación, o bien existe un riesgo inminente de que tales hechos ocurran como consecuencia de haber utilizado el discurso de odio.

Se puede considerar que existe intención de incitar cuando la persona que utiliza el discurso de odio, de forma inequívoca, hace un llamamiento a los demás para que cometan los actos pertinentes o se puede deducir por la contundencia del lenguaje utilizado y otras circunstancias destacables, como la conducta previa del orador. Sin embargo, la intencionalidad no siempre es fácil de demostrar, especialmente cuando las observaciones tienen que ver de forma ostensible con hechos pretendidos o cuando se utiliza un lenguaje codificado.

Es también importante señalar que la Recomendación excluye de forma explícita de la definición de discurso de odio, cualquier forma de expresión, tales como la sátira o informes o análisis realizados de forma objetiva, que simplemente ofenden, dañan o molestan. Al hacerlo, la Recomendación refleja la protección de la definición que adopta el Tribunal Europeo de Derechos Humanos de conformidad con el artículo 10 de la Convención de Derechos Humanos[17]. A pesar de ello, se recuerda que el Tribunal Europeo también reconoce que la incitación al odio puede ser resultado del insulto, la ridiculización o difamación irresponsables de determinados grupos de población, cuya consecuencia puede ser la ofensa innecesaria, la defensa de la discriminación, el uso de un lenguaje vejatorio o humillante o puede incluir la inevitable exposición de la víctima[18] y todas estas formas también estarían incluidas en la definición de la Recomendación.

Por otra parte, para evaluar si existe o no el riesgo de que se produzcan estos actos hay que tener en cuenta las **circunstancias específicas** en las que se utiliza el discurso de odio.

17 Véase por ejemplo Jersild v. Dinamarca [GC], no. 15890/89, 23 de septiembre de 1994, Sürek y Özdemir v. Turquía [GC], no. 23927/94, 8 de julio de 1999, Giniewski v. Francia, no. 64016/00, 31 de enero de 2006, Alves da Silva v. Portugal, no.41665/07, 20 de octubre de 2009 y Fáber v. Hungría, no. 40721/06, 24 de julio de 2012.

18 Vease, por ejemplo, Féret v. Bélgica, no. 15615/07, 16 de julio de 2007 y Vejdeland y Otros v. Suecia, no. 1813/07, 9 de febrero de 2012.

Concretamente, hay que tener en cuenta:

1) El contexto en el que se utiliza el discurso de odio en cuestión (especialmente si ya existen tensiones graves relacionadas con este discurso en la sociedad)
2) La capacidad que tiene la persona que emplea el discurso de odio para ejercer influencia sobre los demás (con motivo de ser, por ejemplo, un líder político, religioso o de una comunidad)
3) La naturaleza y contundencia del lenguaje empleado (si es provocativo y directo, si utiliza información engañosa, difusión de estereotipos negativos y estigmatización, o si es capaz por otros medios de incitar a la comisión de actos de violencia, intimidación, hostilidad o discriminación)
4) El contexto de los comentarios específicos (si son un hecho aislado o reiterado, o si se puede considerar que se equilibra con otras expresiones pronunciadas por la misma persona o por otras, especialmente durante el debate)
5) El medio utilizado (si puede o no provocar una respuesta inmediata de la audiencia como en un acto público en directo)
6) La naturaleza de la audiencia (si tiene o no los medios para o si es propensa o susceptible de mezclarse en actos de violencia, intimidación, hostilidad o discriminación)

RPG No 15 de ECRI

Todos los indicadores son importantes a la hora de identificar y dar seguimiento a un discurso de odio. La organización "Movimiento contra la Intolerancia" ha resaltado la relevancia del conocimiento del significado y mensaje de la simbología racista y neonazi actual de la que se valen aquellos que quieren difundir el odio o la violencia, y para ello buscan referencias al honor, disciplina, valor, espiritualidad, amor a la familia, culto al líder y culto a la guerra. Guarismos como "88" (Heil Hitler), "14NS" (14 palabras de un nacionalsocialista) y RAHOWA (guerra santa racial), por ejemplo son esenciales para indicar e identificar la naturaleza del crimen y de sus autores.[19]

De otro lado, el Manual de Simbología, Comisión Estatal contra la no violencia el racismo, la xenofobia y la intolerancia en el Deporte, es un documento abierto, susceptible de actualizaciones, que recoge un amplio elenco de símbolos, emblemas y banderas, utilizados por distintos grupos, tanto de España como de otros países europeos, y cuya exhibición puede incitar a la violencia, el racismo, la xenofobia o la intolerancia. El objetivo fundamental del Manual es servir de guía tanto a los funcionarios de la Fuerzas y Cuerpos de Seguridad que prestan servicio en los recintos deportivos, como al personal de seguridad privada de los Clubes, para detectar, tanto en las gradas como en los aledaños de los recintos deportivos, a aquellos aficionados que de manera individual o actuando en grupo, puedan portar ese tipo de emblemas, pancartas o banderas.

19 Cuaderno de Análisis Nº 60 "Alerta temprana de los delitos de odio". Movimiento contra la Intolerancia. http://www.educatolerancia.com/wp-content/uploads/2017/11/Cuaderno-de-analisis-60.pdf

3

El enfoque de género

3. El enfoque de género

3.1. La perspectiva de género y la interseccionalidad

El Convenio del Consejo de Europa sobre prevención y lucha contra la violencia contra la mujer y la violencia doméstica[20] incluye el Género en el artículo 3, con las siguientes definiciones:

a) Por «violencia contra la mujer» se deberá entender "una violación de los derechos humanos y una forma de discriminación contra las mujeres, y se designarán todos los actos de violencia basados en el género que implican o pueden implicar para las mujeres daños o sufrimientos de naturaleza física, sexual, psicológica o económica, incluidas las amenazas de realizar dichos actos, la coacción o la privación arbitraria de libertad, en la vida pública o privada".

b) Por «violencia doméstica» se entenderán "todos los actos de violencia física, sexual, psicológica o económica que se producen en la familia o en el hogar o entre cónyuges o parejas de hecho antiguos o actuales, independientemente de que el autor del delito comparta o haya compartido el mismo domicilio que la víctima".

c) Por «género» se entenderán "los papeles, comportamientos, actividades y atribuciones socialmente construidos que una sociedad concreta considera propios de mujeres o de hombres".

El mandato respecto a la igualdad de género y el empoderamiento de las mujeres está acordado por los Estados Miembros y engloba todos los ámbitos de la paz, el desarrollo y los derechos humanos. Los mandatos sobre la igualdad de género toman como base la Carta de las Naciones Unidas que, de manera inequívoca, reafirmó la igualdad de derechos de mujeres y hombres.

20 El Convenio del Consejo de Europa sobre prevención y lucha contra la violencia doméstica, hecho en Estambul el 11 de enero de 2011, ratificado por España y publicado en el BOE el 6 de junio de 2014. Accesible en https://www.boe.es/boe/dias/2014/06/06/pdfs/BOE-A-2014-5947.pdf

La **Cuarta Conferencia Mundial sobre la Mujer** celebrada en 1995 defendió la incorporación de una perspectiva de género como un enfoque fundamental y estratégico para alcanzar los compromisos en igualdad de género. La **Declaración y la Plataforma de Acción de Beijing** resultante instan a tomar medidas en este sentido. Existen compromisos adicionales incluidos en el **documento final del vigésimo tercer periodo extraordinario de sesiones de la Asamblea General**, la **Declaración del Milenio** y diversas resoluciones y decisiones de la Asamblea General de las Naciones Unidas, el Consejo de Seguridad, el Consejo Económico y Social y la Comisión de la Condición Jurídica y Social de la Mujer.

Las **conclusiones del ECOSOC de 1997** definían la incorporación de una perspectiva de género como: "El proceso de evaluación de las consecuencias para las mujeres y los hombres de cualquier actividad planificada, inclusive las leyes, políticas o programas, en todos los sectores y a todos los niveles. Es una estrategia destinada a hacer que las preocupaciones y experiencias de las mujeres, así como de los hombres, sean un elemento integrante de la elaboración, la aplicación, la supervisión y la evaluación de las políticas y los programas en todas las esferas políticas, económicas y sociales, a fin de que las mujeres y los hombres se beneficien por igual y se impida que se perpetúe la desigualdad. El objetivo final es lograr la igualdad [sustantiva] entre los géneros".

El género es un operador social central en la configuración de las jerarquías sociales, pero no opera de forma autónoma sino intersectado o en combinación con otros operadores sociales, principalmente la clase social y la etnia. La herramienta conceptual que nos permite visibilizar la interceptación de los diferentes operadores sociales (género, clase, etnia, etc.) se denomina interseccionalidad. Este enfoque interrelaciona distintas categorías que participan en la formación de la identidad, construidas social, política, económica, cultural y psicológicamente, dando lugar a posiciones diferenciadas entre unas personas y otras en la sociedad (Helia del Rosario Rodríguez, Derecho a una vida libre de violencias. Experiencias y resistencias desde las mujeres migrantes: Estudio de casos).

3.2. El enfoque de género en el Protocolo e Indicadores del proyecto ALRECO

La consideración del principio de igualdad de trato y no discriminación entre hombres y mujeres es un factor esencial en el proyecto ALRECO, ya que las mujeres constituyen uno de los grupos más vulnerables y por tanto con mayor riesgo de ser el objetivo de incidentes y discursos racistas o xenófobos.

Cuando al hecho de la discriminación por origen racial o étnico se le añade el hecho de ser mujer se hace evidente el doble peso de la discriminación por motivo de género, origen étnico y otras formas conexas de intolerancia. Las desventajas que encaran las mujeres pertenecientes a minorías en relación con el mercado de trabajo, la trata y la violencia basada en la raza, constituyen esferas de especial preocupación. Este contexto es el que sustenta la desigualdad múltiple sufrida por las mujeres en razón de su género, su pertenencia a esta o aquella etnia y/o raza y las posibilidades reales de hacer frente a sus necesidades. Según las Naciones Unidas para muchas mujeres, los factores relacionados con su identidad social, como la raza, el color, el origen étnico y el origen nacional se convierten en diferencias que pueden crear problemas que afectan sólo a grupos particulares de mujer o que afectan a algunas mujeres de manera desproporcionada en comparación con otras y a sus posibilidades de hacer frente a sus necesidades.

En los últimos años el discurso de odio racista y xenófobo ha tratado, a menudo, de vincular la inmigración con la violencia de género. Relacionar el maltrato hacia las mujeres con la inmigración, pretende estigmatizar a determinados grupos relacionándolos con la violencia, el machismo, etc. Los informes alertan del incremento de odio con graves agresiones a mujeres musulmanas, en lo que se ha denominado la Islamofobia de Género[21].

Otro ejemplo que cabe señalar en el discurso de odio e intolerancia que conceptualiza despectivamente la categoría

21 Plataforma Ciudadana contra la Islamofobia. Accesible en http://www.observatorioislamofobia.org/que-es-la-islamofobia/

analítica “género” como una ideología es el insulto “feminazi”, que pretende justificar las discriminaciones y las prácticas nocivas contra las mujeres, principalmente, a través de juicios morales sobre su vida, formas de vestir, ideología, etc. profundizando en un discurso de odio que refuerza la misoginia, la intolerancia y que justifica la violencia verbal y física contra las mujeres.

En este contexto, el proyecto ALRECO incluye en su Protocolo y Sistema de Indicadores para detectar el discurso de odio en las redes sociales, indicadores que tengan en cuenta el factor género como un elemento identificativo y/o agravante del discurso de odio, centrando alertas, principalmente en la detección de:

- Racismo/xenofobia de género.
- Vinculación entre migración y violencia de género.
- Islamofobia de género.

3.3. El enfoque basado en los derechos humanos

La universalidad de los derechos iguales e inalienables de todos los seres humanos establecen las bases para la libertad, la justicia y la paz en el mundo, según la Declaración Universal de Derechos Humanos, adoptada por la Asamblea General de las Naciones Unidas en 1948, fundamentados en la dignidad de las personas y su libertad. La prioridad de aplicar los principios de los derechos humanos fue la piedra angular de las iniciativas de reforma de las Naciones Unidas que comenzaron en 1997.

El enfoque basado en los derechos humanos se centra en las personas vulnerables y los grupos de población que son objeto de una mayor marginación, exclusión y discriminación. y sufran unos niveles más elevados de intolerancia, discurso y delitos de odio. Este enfoque, a menudo, requiere un análisis de las normas de género, de las diferentes formas de discriminación y de los desequilibrios de poder a fin de garantizar que las intervenciones lleguen a los segmentos más oprimidos y segregados de la población.

Como se señaló anteriormente, la tipificación jurídica de lo que se ha venido a denominar delitos de odio, afecta a cualquier infracción penal sometida a la circunstancia agravante del art.22.4 del Código Penal español, y junto al denominado discurso de odio, a los artículos del Capítulo IV: **De los delitos relativos al ejercicio de los derechos fundamentales y libertades públicas**, bien sea, delitos cometidos con ocasión del ejercicio de los derechos fundamentales y de las libertades públicas garantizados por la Constitución, y los delitos contra la libertad de conciencia, los sentimientos religiosos y el respeto a los difuntos. Esta tipificación se corresponde con la protección de los Derechos Fundamentales, de españoles y extranjeros , de sus libertades públicas, en la Constitución Española y significativamente del artículo 10: 1. La dignidad de la persona, los derechos inviolables que le son inherentes, el libre desarrollo de la personalidad, el respeto a la ley y a los derechos de los demás son fundamento del orden político y de la paz social. 2. Las normas relativas a los derechos fundamentales y a las libertades que la Constitución reconoce se interpretarán de conformidad con la Declaración Universal de Derechos Humanos y los tratados y acuerdos internacionales sobre las mismas materias ratificados por España.

En consecuencia, se consideran elementos de buenas prácticas propios del enfoque basado en los derechos humanos, entre otros:

- las actividades que ven en el pleno ejercicio de los derechos humanos el fin último de su desarrollo;
- la participación de las personas que es a la vez un medio y un objetivo;
- las estrategias que proporcionan fortalecimiento y autonomía, en lugar de negarlo;
- el análisis de situación se utiliza para identificar las causas inmediatas, subyacentes y fundamentales de los problemas de violación de derechos humanos, por discriminación y odio basado en la intolerancia.

El análisis incluye a todos los grupos de interés, entre ellos, las capacidades del Estado como principal garante de derechos y el papel de otros agentes no estatales como las ONGs y otros. En este sentido, habría que destacar una doble necesidad: en primer término, la protección de las personas y de las comunidades o colectivos sociales, que requieren que se las informe acerca de sus derechos, y en segundo lugar, su protección de los discursos y delitos de odio.

METODOLOGÍA

4

Metodología

4. Metodología

Dentro 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. 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.

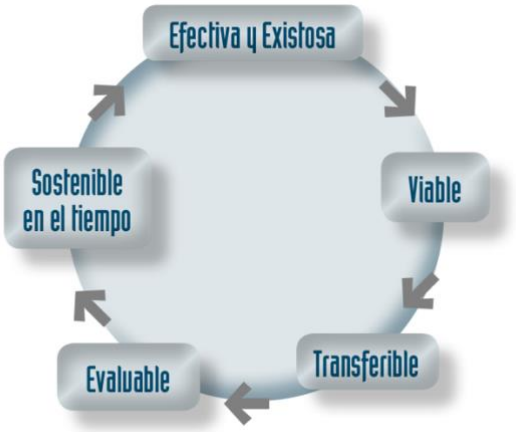
En el marco de esta acción se contempla como paso previo: la identificación de experiencias y buenas prácticas que se hayan desarrollado en la Unión Europea que sirvan de base para el debate y tipificación de indicadores de alerta temprana.

A continuación, resumimos la metodología que se empleó para la selección de buenas prácticas y para el desarrollo del protocolo y sistema de indicadores.

4.1. Metodología para la selección de las buenas prácticas

Tomando como punto de partida el hecho de que una buena práctica no es tan sólo una práctica que se define como buena en sí misma, sino que es una práctica que se ha demostrado que funciona bien y produce buenos resultados, y, por lo tanto, se recomienda como modelo. Se trata de una experiencia exitosa, que ha sido probada y validada, en un sentido amplio, que se ha repetido y que merece ser compartida con el fin de ser adoptada por el mayor número posible de personas. En este sentido deberá cumplir al menos 5 criterios:

Figura 1



BUENA PRÁCTICA

Aquella actuación, metodología o herramienta desarrollada 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, antisemitas y antigitanos.

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, disponga de mecanismos de evaluación, heterogeneidad del conjunto de experiencias en cuanto al agente promotor y los beneficiarios de la experiencia (institucional, universidad, impulsadas por el tercer sector, etc.) y que cuente con mecanismos de coordinación.

Teniendo en cuenta estas consideraciones, se ha consensuado entre los socios del proyecto ALRECO una definición de lo que se entenderá como buena práctica en relación con las herramientas y experiencias en el ámbito del discurso de odio en línea.

En segundo lugar, para la identificación de buenas prácticas se han implementado diferentes herramientas metodológicas:

Tabla 1

HERRAMIENTAS	INSTRUMENTOS
ANÁLISIS DOCUMENTAL	Documentos principales: • Proyectos similares • Herramientas ya existentes • Experiencias previas • Artículos académicos e 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), y la 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 de sistematización de la información para la valoración de las buenas prácticas.

En tercer lugar se ha procedido a contactar con diferentes personas y entidades clave para tener un mayor alcance en la búsqueda de buenas prácticas:

Tabla 2

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
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

Con la metodología empleada se ha obtenido información de un total de 53 experiencias en el periodo comprendido entre el 9 y el 30 de enero de 2019. De las cuales se han seleccionado 18, por ser las que cumplían los criterios establecidos. 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" para la detección de discurso de odio si no que, más bien, tienen una parte de herramienta, otra parte de contranarrativa, algunas de sensibilización, de formación. De las 18 experiencias seleccionadas, solo 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 objeto 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", lo que también 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 la que es efectiva (p.ej no tenemos ninguna herramienta de países árabes pero sí que trabajen en ese idioma). En definitiva, disponemos de experiencias de los siguientes países: España, Francia, Italia, Reino Unido, Grecia, Alemania, Estados Unidos.

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

En cuanto al tipo de promotores de las experiencias seleccionadas, básicamente son tres: universidades/ámbito académico, instituciones públicas y ONG´s. La mayoría de las experiencias se financiaron a través de fondos públicos, si bien algunas fueron financiadas por empresas como Google o Facebook.

4.2. Metodología para el desarrollo del protocolo y del sistema de indicadores

De la revisión aportada en el primer informe de buenas prácticas, así como de la revisión de artículos específicos que profundizan en herramientas existentes sobre monitorización del discurso de odio, se han seleccionado 16 herramientas. Esta selección nos ha permitido profundizar en la metodología. La siguiente tabla muestra el grado de relevancia de cada una de las buenas prácticas, específicamente de sus metodologías, con el fin último de crear una herramienta propia (tabla 3).

Tabla 3
Herramientas para identificar discurso de odio en la red

HERRAMIENTA	RELEVANCIA DE LA METODOLOGÍA PARA **ALRECO**	OTROS COMENTARIOS
Clasificador de OBERAXE	Alta	Tipología de clasificación.
Somos más (Proyecto de Google)	Baja	No tiene una herramienta de monitorización. Proyecto centrado en la sensibilización, intervención y en la promoción del ciberactivismo para contrarrestar el discurso de odio y radicalismo. Especialmente centrado en YouTube.
Ciber hache	Media	No tiene una herramienta de monitorización. Captación de tweets a través de hashtag-trending topics. Interesante la técnica de análisis utilizada a través de la minería de datos. Análisis por pares de expertos y contraste con Kappa test.

HERRAMIENTA	RELEVANCIA DE LA METODOLOGÍA PARA ALRECO	OTROS COMENTARIOS
CibeRespect	Media	Proyecto centrado en la intervención y promoción del ciberactivismo contra el discurso de odio. Se hace un seguimiento del discurso de odio. No obstante, no tiene herramienta de monitorización.
Observatorio PROXI	Alta	Monitoreo de noticias sobre población migrante y población gitana de 3 medios digitales con gran audiencia en España. Actualmente, no está en funcionamiento el equipo de análisis. El software está disponible en acceso abierto. Propuesta de ciberactivismo.
Be the Key	Baja	Proyecto centrado en la intervención y promoción del ciberactivismo. Especialmente, centrado en la islamofobia. La plataforma principal de difusión es Facebook.
Official rewind	Baja	Proyecto centrado en la intervención para promover la contranarrativa, a través de Twitter y Facebook.
Getthetrollsout	Baja	Monitoreo cualitativo de redes sociales y media. Proyecto orientado a la intervención. Especial atención al discurso de odio contra minorías religiosas.
MANDOLA	Alta	Creación de una App para reportar discurso de odio en la red. Basado en algoritmo para medir discurso de odio producido por país. Existe un mapa comparativo de países.
Wordsarestones	Baja	Proyecto centrado en la intervención a través de la formación en la detección del discurso de odio, y la promoción del ciberactivismo.

HERRAMIENTA	RELEVANCIA DE LA METODOLOGÍA PARA **ALRECO**	OTROS COMENTARIOS
Save a hater	Baja	Proyecto centrado en la intervención para crear contranarrativas a través del ciberactivismo.
Hate meter	Media	Proyecto en desarrollo que pretende construir una herramienta. Centrado en islamofobia.
Silence hate	Baja	Proyecto centrado en la intervención
Contra l'odio	Alta	Tiene un mapa con datos sobre el tipo y frecuencia de discurso de odio en cada zona de Italia. Se basa en una herramienta para el seguimiento en Twitter a través de un algoritmo para favorecer el aprendizaje automático y procesamiento de lenguaje natural en la identificación automática en redes sociales. El objetivo será un "score" de la cuenta de Twitter que permita identificar la tendencia a usar lenguaje de odio o a seguir personas que lo usan.
Hate base	Alta	Análisis lingüístico de conversaciones públicas para derivar una probabilidad de contexto de odio. Basado en natural language engine (Hate brain) algoritmos. Utilización de un amplio vocabulario
Donate the hate	Baja	Proyecto centrado en la formación, sensibilización e intervención contra el discurso de odio.

HERRAMIENTA	RELEVANCIA DE LA METODOLOGÍA PARA ALRECO	OTROS COMENTARIOS
Exploring Online Hate	Media	Análisis de usuarios específicos productores de discurso de odio. Identificación de hashtags y términos más usados. Data mining y creación de algoritmos, pero no es información pública.
Saferlab		Uso de métodos de qualitative and natural language para identificar mecanismos de la violencia y discurso de odio.
Smartphone application (App) called "LIGHT ON RACISM"	Baja	Promover el ciberactivismo contra el discurso de odio. Herramienta para promover la denuncia y la contranarrativa.

Se ha profundizado en la propia web de las herramientas, y también en artículos y documentos sobre la misma. De cada herramienta se han analizado los siguientes aspectos:

- Metodología de la herramienta (cómo funciona)
- Tipos de lenguaje especificados en la herramienta (categorías establecidas respecto al lenguaje del discurso de odio, si las tienen. Posibles indicadores de las mismas)
- Tipos de objetivo/finalidad del discurso (categorías establecidas por la herramienta respecto a los objetivos o finalidades del discurso de odio, si las tienen. Posibles indicadores de las mismas)
- Tipos de intensidad del discurso de odio (categorías establecidas por la herramienta respecto a la intensidad, si las tienen. Posibles indicadores de las mismas)

Para complementar estas informaciones, se ha contactado con los grupos responsables de algunas de las herramientas, concretamente PROXI desarrollada por el IDHC (Institut de Drets Humans de Catalunya) o con Cyberhache.

Adicionalmente se han consultado artículos científicos a través de diferentes bases de datos científicos, como el buscador ISI Web of Science, con el objetivo de recoger las claves metodológicas, desarrollos y retos actuales.

Los términos utilizados para la búsqueda han sido los siguientes:

Key words Web of Science (últimos 10 años)
Hate Speech and Social Media
Hate Speech and Social Networks
Hate Speech Detection
Cyberhate and Social Networks.

Después, para concretar la búsqueda hacia aquellos colectivos que atiende la investigación, se emplearon los siguientes términos: Hate Speech and Racism, Hate Speech and Xenophobia, Hate Speech and Immigration, Hate Speech and Islamophobia, Hate Speech and muslims, Hate Speech and Antisemitism, Hate Speech and Antisionism.

De todo el análisis realizado se seleccionaron 18 artículos útiles para el desarrollo de nuestra herramienta.
De los artículos seleccionados, se analizó la información que aportan sobre todos o algunos de los siguientes aspectos:

- Formas de análisis del discurso de odio.
- Aspectos metodológicos (descripción y clasificación de metodologías, revisión de ventajas e inconvenientes de diferentes metodologías, etc).
- Herramientas desarrolladas específicas. (En ese caso, objetivo y alcance de la iniciativa, impacto logrado, evaluación de la misma).
- Tipificaciones del discurso de odio (semánticas, tipificaciones de contenido, intensidad, otros).
- Aportaciones relacionadas con las diferentes áreas de AL-RE-CO (racismo, xenofobia, antisemitismo, islamofobia) y con la perspectiva de género de forma transversal.
- Otros aspectos de interés.

El desarrollo de este proceso metodológico permite extraer las siguientes conclusiones:

- Existen diferentes experiencias, tanto herramientas como artículos, que son de gran utilidad para el desarrollo de la herramienta prevista en el proyecto ALRECO.
- Es importante ser conscientes de las limitaciones intrínsecas a una herramienta basada en un algoritmo. Las experiencias y artículos analizados hacen especial mención a estos límites.
- Es necesario acotar al máximo el alcance de la herramienta en diferentes aspectos: términos de búsqueda, idioma en el que se va a desarrollar (y zonas geográficas), ámbitos concretos (racismo, xenofobia, antisemitismo, islamofobia y antigitanismo) y con la perspectiva de género de forma transversal).

BANCO DE PALABRAS E INDICADORES

5

Punto de partida

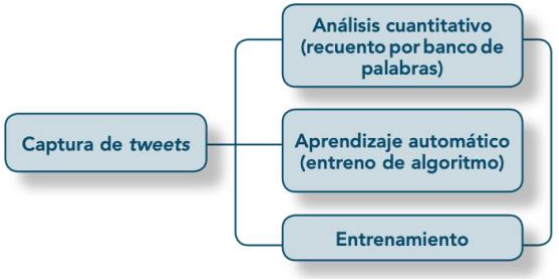
5. Punto de partida

La propuesta de protocolo se ha realizado a partir de la experiencia del equipo en la creación de metodologías de análisis del discurso en las redes sociales (Cabré-Olivé, Flecha-García, Ionescu, Pulido, & Sordé-Martí, 2017; Pulido, Redondo-Sama, Sordé-Martí, & Flecha, 2018) así como del marco teórico sobre discurso de odio, literatura especializada en la detección informática de discurso de odio en las redes, y la exploración de buenas prácticas.

Las principales aportaciones teóricas sobre análisis del discurso y racismo (Van Dijk, 2002, 2007; Wieviorka, 1992) y sobre los actos comunicativos (Austin, 1961; Searle & Soler, 2004) señalan la complejidad del racismo y odio en el discurso, y su dependencia del contexto y de elementos no verbales. Por su lado, los estudios sobre discurso de odio desde el punto de vista de las ciencias informáticas son todavía escasos (Fortuna y Nunes, 2018). Los artículos consultados coinciden en señalar las dificultades para identificar y clasificar el discurso de odio en internet, en concreto (Hughey & Daniels, 2013): importancia del contexto y emisor para determinar el sentido de odio o no de una misma palabra o expresión; expresiones no explícitas de racismo (sutilezas, metáforas, etc); especificidad de discurso según los diferentes grupos objetivos; interseccionalidad; volumen de datos.

La mayoría de experiencias existentes utilizan como características para detectar discurso de odio un conjunto de palabras o "bag of words" (Gitari, Zuping, Damien, & Long, 2015; Greevy & Smeaton, 2004). Como hay obvias limitaciones, en muchas ocasiones se intenta la combinación de diferentes palabras o partes de texto. Por otro lado, diversas investigaciones están tratando de mejorar la automatización de la detección del discurso de odio mediante machine Learning o aprendizaje automático. Es decir, mediante algoritmos que mejoran con el uso o entrenamiento de la detección de discurso de odio. El aprendizaje automático incluye una primera etapa de recolección de los tweets (con diferentes criterios) a través de un listado de palabras definidas (Burnap & Williams, 2015, 2016). La segunda etapa consiste en el entrenamiento de un algoritmo, habitualmente a partir de la participación de personas como codificadoras de los textos.

El protocolo que presentamos parte del siguiente esquema:



A continuación se exponen las estrategias de captura y de análisis de contenido.

6

Captura de tweets

6. Captura de tweets

Para intentar acotar el monitoreo de discurso de odio a nivel nacional - no existe una manera mediante Twitter de poder saber si un tweet procede o no de España - se capturarán tweets en lengua castellana, y se seleccionarán las franjas horarias donde se publican mayoritariamente los tweets en España (para diferenciar los tweets procedentes de Latinoamérica).

Dado el volumen de tweets diarios, se han definido tres estrategias de captura de tweets para obtener muestras manejables sobre las que analizar el discurso de odio.

6.1. Bottom Up: Captura de tweets por temas importantes del día

La estrategia Bottom Up pretende facilitar la identificación de discurso de odio por los temas populares en Twitter (Trending topics) según sus usuarios, así como su posible relación con sucesos u otros factores que contribuyan a su incremento.

El proceso para la captura Bottom Up es:

1. Programar la obtención automática diaria de los 50 temas más importantes del día (España).
2. Una vez obtenida la lista de los 50 temas, se programará la obtención automática de todos los tweets publicados bajo cada uno de los temas, por tanto, se obtendrán 50 listas de tweets completos correspondientes a cada uno de los temas.

La ejecución de la rutina de captura y de análisis de los temas más importantes del día y la obtención de los tweets para cada tema, se ejecutará cada día a la misma hora especificando la fecha correspondiente.

Ejemplo captura realizada día 17/04/2019:

#DiAlgoBonitoAUnaEscritora	Tour de la Manada	PSOE y Compromís	#STOPOKUPAS	#DebateTVE
#FelizMiércoles	España y Turquía	Feliz Miércoles Santo	#AR17A	#GabrielGarciaMarquez
#DebateESP	William Carvalho	El Consejo de Europa	#DiaMundialdelaHemofilia	#DíaMundialHemofilia
#MiércolesSanto	Arrimadas a Rufián	Premio Nobel de Literatura	#LaCafeteraJEcJEc	#17Abril
#ApodérateUnidasPodemos	Televisión Española	Open Arms	#RumboUrnasARV	#CasoGrúas
Atresmedia	Gabo	Damian Lillard	#STRP	#WednesdayMotivation
Gabriel García Márquez	cartaya	Junta Electoral a Vox	#Homecoming	#DíadelaInformaciónJuvenil
RTVE	El Hijo	McCollum y Lillard	#NoEsNo	#SuperSmashBrosUltimate
Manuel Alcántara	Mt 26	Pedro Saura	#BeyoncéHomecoming	#SoloSíEsSí
Chavela Vargas	Juanma Lillo	#YoigoTeDaUnPlus	#hemofilia	#HdadNervión19

El objetivo es capturar los *tweets* que hagan referencia explícita a las comunidades objeto de discurso de odio seleccionadas en el proyecto, para identificar la presencia y tipos de discurso de odio hacia las mismas. El punto de partida serán palabras que se utilizan habitualmente para hacer mención de estas comunidades.

El listado inicial de palabras para la captura de *tweets* es el de la tabla 4. Se detectarán los *tweets* que contengan las palabras que se listan en la tabla. No obstante, es relevante señalar que los *tweets* que contengan estas palabras no sirven para determinar la existencia de discurso de odio. Es un listado que solo sirve para capturar *tweets* relacionados con colectivos que están en el análisis.

6.2. Top Down: Captura de tweets referentes a los colectivos

Tabla 4
Palabras para la captura de tweets (provisional)

Colectivo	Términos
Personas migrantes y extranjeras	migracion inmigracion migrante migrantes inmigrante inmigrantes extranjero extranjera extranjeros extranjeras refugiado refugiada refugiados refugiadas marroqui marroquis magrebi magrebis africano africana africanos africanas latino latina latinos latinas rumano rumana rumanos rumanas subsahariano subsahariana subsaharianos subsaharianas sudamericano sudamericana sudamericanos sudamericanas sudaca sudacas chino china chinos chinas chinito chinita clandestino clandestinos mena menas
Personas de minorías culturales, étnicas o religiosas	negro negra negros negras* panchito panchita panchitos panchitas indio india indios indias conguito conguita conguitos conguitas machupichu machupichus
Personas gitanas	gitano gitana gitanos romani romanis patriarca
Personas musulmanas	musulman musulmana musulmanes musulmanas moro mora moros moras islamico islam arabe arabes sarraceno sarracenos yijadista yijadistas islamista islamistas fundamentalista fundamentalistas fundamentalismo hiyab burca
Personas judías	judio judia judios judias sion sionista sionistas hebreo israeli israelis holocuento holocausto hitler camara+gas nazi
Globales	stopinvasion mantero manteros topmanta eurabia terrorismo atentado atentados guerra sumision boicot frontera fronteras expulsion

Este es un listado de palabras provisional, que se mejorará a partir del pilotaje de la herramienta. La captura de tweets por palabras claves predefinidas se realizará semanalmente. Si el volumen de datos es excesivo se realizarán capturas diarias.

6.3. Top Down: Captura de tweets por criterios específicos

Otra estrategia para capturar tweets es personalizando la captura a partir de usuarios, hashtags, temas o eventos, donde por algún motivo se haya detectado que se está generando o se puede estar generando discurso de odio. Por ejemplo: un atentado terrorista, una persona pública, una noticia de sociedad. La ejecución de estas rutinas se hará de forma manual y en función de necesidades o de oportunidad según acontecimientos.

BANCO DE PALABRAS
E INDICADORES

7

Análisis del discurso de odio

PROTOCOLO
Y SISTEMA DE INDICADORES
para la detección del discurso de odio en las redes sociales

7. Análisis del discurso de odio

Para cada una de las muestras de tweets obtenidas se procederá a tres tipos de análisis del discurso de odio.

7.1. Análisis cuantitativo por banco de palabras

Esta estrategia permite clasificar el contenido de los tweets en Discurso de odio y No discurso de odio, en función de si contienen palabras identificadas como propias de discurso de odio (banco de palabras). Se propone un Banco de palabras propio con dos categorías de términos (a los que habrá que añadir el plural y el femenino de algunas de ellas).

Este banco de palabras surge de la exploración de 5 bancos existentes, un grupo de discusión, un análisis cuantitativo del léxico de 20 perfiles racistas en twitter, y análisis cualitativos complementarios de tweets (más de 1000) identificados como discurso de odio. En el anexo II se detalla la metodología seguida.

El banco de palabras es necesariamente limitado y no puede recoger la diversidad de términos utilizados en las diferentes expresiones de discurso de odio. Aunque sea imperfecto, sin embargo, supone un punto de partida para el diseño y pilotaje de la herramienta de monitoreo del discurso de odio. Se trata por lo tanto de un banco de palabras que debe quedar abierto a modificaciones (incorporaciones, matices, eliminaciones) durante la fase del pilotaje. Además, una vez en funcionamiento, se propone que el banco de palabras sea de acceso abierto y que la ciudadanía pueda aportar nuevas palabras y comentarios, siendo así que la ciudadanía tenga derecho de intervenir no sólo en la identificación del discurso de odio, sino también en la propia herramienta.

7.2. Análisis cuantitativo con aprendizaje automático

Esta estrategia pretende identificar patrones y relaciones entre las palabras del texto que permitan clasificar el contenido de los tweets según diferentes tipos de intensidad de odio (de odio extremo a discurso upstander).

Se utilizarán técnicas de machine learning supervisado para la implementación y entrenamiento de un algoritmo. Ello requiere la siguiente secuencia:

1. **Limpiar los datos**: Se eliminan símbolos o caracteres que no aportan significado al mensaje de los tweets, separar las palabras y homogeneizar las palabras para poner el foco en el contenido y no en la forma (convertir los caracteres minúsculas, abreviaciones, etc).

2. **Etiquetaje de los tweets para el entrenamiento**: El conjunto de tweets se analizará de forma manual, haciendo análisis del contenido de los tweets. Por lo tanto, se requiere de personas revisoras de tweets, y cada persona revisora etiquetará los tweets analizados según una escala de intensidad:

Odio extremo	Discurso que incita a la violencia.
Odio-ofensa	Discurso que representa ofensas personales o colectivas, que incita a la discriminación, reproduce tópicos y falsedades.
Discurso neutro	Discurso descriptivo, en el que no aparece odio.
Upstander	Discurso alternativo, que contribuye a una contra-narrativa, rompiendo con los tópicos o posicionándose en defensa de los colectivos objeto de odio.

Elaboración propia a partir de fuentes: Watanabe, Bouazizi, & Ohtsuki, 2018; Gitari, Zuping, Damien, & Long, 2015; Hate base

Para el análisis de contenido, cada revisor/a tendrá codebook (definición de los códigos) de antemano. En el entrenamiento, se valorará la posibilidad de examinar el nivel de acuerdo en la asignación de las categorías (por ejemplo, coeficiente de concordancia Kappa).

3. **Obtención del modelo**: Con los tweets etiquetados manualmente se creará un modelo, a partir del cual el algoritmo será capaz de clasificar los tweets según la escala de intensidad de discurso de odio definida. Este proceso requiere ajustar diferentes modelos y seleccionar cuál es el que mejor responde a la clasificación deseada.

4. **Clasificación o predicción**: Finalmente a partir de la construcción del modelo de predicción, se aplicará a los tweets no clasificados para clasificarlos según su tipología.

Las diferentes fases que componen la implementación del algoritmo de análisis, suponen un proceso reiterativo de ajuste, basado en analizar los resultados obtenidos y de ir afinando los parámetros de ajuste y selección.

El modelo obtenido se irá actualizando de forma periódica incorporando nuevos conjuntos de tweets clasificados como discurso de odio para ir ajustando el modelo.

7.3 Análisis cualitativo manual

Además de los análisis cuantitativos mencionados, la herramienta posibilitará el análisis cualitativo de tweets de forma manual (no automatizada), como una opción voluntaria para usuarios/as que quieran profundizar en el análisis que los algoritmos y la automatización no permiten.

Para ello, se facilitará un documento de pautas y orientaciones con ejemplos, centrándose en diferentes características: tipos de lenguaje finalidad del lenguaje, registros lingüísticos.

Tipos de lenguaje utilizado en el discurso de odio

- Lenguaje insultante y degradante
- que incita o realiza apología de la violencia
- Justificación, bromas, trivialización de la violencia hacia 'los otros'
- Lenguaje divisorio o de otredad (ellos versus nosotros)
- Estereotipo-Prejuicio
- Rumores
- Hechos falsos (False Facts)
- Argumentos Trampa (Flawed argumentation)
- Metáforas-comentarios deshumanizadores-ironías

Fuentes de referencia: Noriega & Iribarren, 2012; Observatorio Proxi, 2015; Van Dijk, 2002, Cyberhache, Contro l'Odio

Tipos de "otredad"

Como señala la literatura científica, el discurso de odio se caracteriza por una polarización muy marcada entre el Ellos y el Nosotros, donde para Ellos se categorizan y generalizan cualidades negativas en contraposición de cualidades positivas del propio grupo. La transmisión del mensaje puede dirigirse a los otros, interpelándolos directamente, o puede hacer referencia A los otros sin dirigirse a ellos.

- Lenguaje dirigido a los otros
- Lenguaje sobre los otros

Tipos de finalidad del discurso de odio

- Ofender
- Ofensa personal
- Ofensa colectiva
- Incitar a la violencia
- Incitación a la discriminación
- Incitación a la segregación

Fuentes de referencia: Miró Llinares, 2016; Cyberhache, Proxi, Miró.

Tipos de registros lingüísticos

- Estructuras no-verbales despectivas. Ej. Emoticonos, puntuación, mayúsculas, etc.
- Sintaxis. Ej. Enfatizar o desenfatizar la acción, a través de oraciones activas vs. pasivas
- Léxico. Ej. Terrorista vs. luchador por la libertad
- Significado local de una oración
- Significado vago o indirecto de "nuestra" acción vs. significado detallado sobre "sus" conductas impropias
- Significado global del discurso. Ej. Temas abordados positivos para nosotros (solidaridad, tolerancia, etc.)/ negativos para ellos (crimen, violencia, etc.)
- Esquemas. Ej. Simplificación
- Dispositivos retóricos, metáfora, metonimia, hipérbole, eufemismo, etc.
- Interacción. Interrumpir, terminar antes, discrepar agresivamente, no responder

Fuentes de referencia: (Van Dijk, 2002, 2007)

Tipos de discurso Upstander

El discurso Upstander se caracteriza por la proactividad en la denuncia y/o la aportación de discurso positivo relacionado con los colectivos víctimas del discurso de odio. Los contenidos pueden clasificarse de la siguiente manera:

- Publicación de comentarios propios: argumentación con evidencias desmintiendo prejuicios e visibilización de experiencias positivas.
- Intervenciones antirumor: datos y explicaciones contra-rumor.
- Intervenciones pedagógicas: hechos y datos para clarificar el debate.
- Intervenciones sensibilización: apelar a las emociones.
- Cualificación de comentarios de otros usuarios: valorar positivamente o descalificar comentarios positivos/negativos.
- Denuncia de comentarios de odio: a través de las herramientas propias de las redes sociales y apps creadas específicamente para ello.

Fuentes de referencia: Observatorio Proxi, 2015

Las definiciones y categorías son un punto de partida, en la fase de desarrollo y pilotaje de la herramienta se modificarán en función de los resultados obtenidos.

Además del uso para el análisis cualitativo, la herramienta debería facilitar pautas para la intervención activa en las "contranarrativas".

8

Sistemas de indicadores

8. Sistema de indicadores

Distinguimos entre los indicadores que permiten identificar un *tweet* como discurso de odio (y los diferentes tipos de intensidad) y los indicadores de discurso de odio que se podrán obtener de la implementación de la herramienta de monitoreo del discurso de odio.

Indicadores de discurso de odio en las redes sociales

Las siguientes características son indicadores de discurso de odio que en el desarrollo de la herramienta, se utilizarán para la fase de pilotaje y de entreno del algoritmo (comprobar si los tweets están correctamente identificados como discurso de odio, etiquetar los tweets según su intensidad).

Tipos de discurso de odio

- Presencia de palabras (Banco de palabras)
- Lenguaje que incita o realiza apología de la violencia
- Justificación, bromas, trivialización de la violencia hacia 'los otros'
- Lenguaje divisorio o de otredad (Ellos versus Nosotros)
- Reproducción de estereotipos o prejuicios
- Difusión de rumores
- Difusión de datos falsos (False Facts)
- Argumentos Trampa (Flawed argumentation)
- Metáforas, comentarios deshumanizadores, ironías
- Estructuras no-verbales despectivas. Ej. Emoticonos, puntuación, mayúsculas, etc.

Con estas características, el discurso de odio puede distinguirse entre extremo u ofensivo.

- **Extremo**: incitación a la violencia.
- **Ofensivo**: ofensas personales o colectivas, que incita a la discriminación, reproduce tópicos y falsedades.

Indicadores para el monitoreo del discurso de odio

La implementación de la herramienta permitirá analizar periódicamente el discurso de odio en las redes sociales. Se generarán diversos indicadores para monitorizar la evolución temporal del discurso de odio, la intensidad del discurso de odio, y otras características.

Tabla 6
Indicadores de presencia de discurso de odio

Indicador	Descripción y variaciones
Num. de tweets con discurso de odio	Número de tweets con presencia de palabras de odio (banco de palabras). Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un Evento/tema específico
Porcentaje de tweets con discurso de odio	Porcentaje de tweets con presencia de palabras de odio respecto a los tweets que no contienen estas palabras. Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un evento/tema específico
Porcentaje de discurso de odio en los temas del día	Porcentaje de temas del día en los que hay presencia de tweets con las palabras de odio.
Crecimiento del discurso de odio en los tweets sobre los colectivos	Porcentaje de crecimiento del discurso de odio entre dos fechas determinadas. Aplicable a: • Número total de tweets con discurso de odio • Porcentajes de tweets con discurso de odio

Observaciones: La obtención de estos indicadores requiere de un banco de palabras lo más idóneo posible. Como ya se ha comentado, estos indicadores no recogerán todo el discurso de odio sino sólo el que utiliza palabras claramente identificables o potencialmente de discurso de odio. Tampoco podrán recoger el discurso de odio existente que no utiliza las palabras del banco de palabras. Sin embargo, aportan información relevante sobre la evolución de un tipo de discurso de odio explícito.

Tabla 7

Indicadores de intensidad de discurso de odio

Indicador	Descripción
Num. de tweets clasificados como odio extremo	Número de tweets que incitan a la violencia. Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un evento/tema específico
Num. de tweets clasificados como odio - ofensa	Número de tweets que representan ofensas personales o colectivas, que incita a la discriminación, reproduce tópicos y falsedades. Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un evento/tema específico
Num. de tweets clasificados como neutros	Número de tweets que no contienen discurso de odio ni tampoco una contranarrativa. Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un evento/tema específico
Num. de tweets clasificados como upstander	Número de tweets con discurso alternativo, que contribuye a una contra-narrativa, rompiendo con los tópicos o posicionándose en defensa de los colectivos objeto de odio. Aplicable a: • Cada uno de los temas importantes en un día • Conjunto de los temas importantes en un día • Cada uno de los colectivos analizados durante un tiempo determinado • Conjunto de tweets relacionados con un evento/tema específico

Observaciones:

La obtención de estos indicadores requiere del entrenamiento de un algoritmo con la participación de personas que etiqueten tweets. La alta complejidad del discurso racista, como ya se ha comentado, condiciona las posibilidades de lograr un modelo eficaz para la identificación de diferentes grados de intensidad.

Si el modelo permite obtener alguna o todas estas categorías, pueden desarrollarse otros indicadores, como porcentajes y tendencias temporales.

9

Especificaciones técnicas

9. Especificaciones técnicas

Las tres estrategias de recogida de datos se implementarán con una aplicación desarrollada con lenguaje de programación Python para realizar las búsquedas y capturas correspondientes.

Usuarios, roles ylimitaciones

La aplicación de monitorización de discurso de odio y la visualización de sus resultados requiere de procesar un gran volumen de información. Los recursos informáticos en relación al almacenamiento, la memoria y la potencia de cálculo son críticos y, por tanto, se debe controlar el consumo de recursos en cada momento. Por este motivo, es necesario limitar la ejecución de los algoritmos en periodos temporales concretos, así como el número de usuarios y el uso de las funcionalidades disponibles para cada uno de ellos.

La propuesta inicial de usuarios y roles será la siguiente:

Administrador/a	Configurar los parámetros de captura de tweets. Configurar el algoritmo de detección discurso de odio. Configurar indicadores y estadísticos para el discurso de odio. Configurar los parámetros para la visualización de datos y filtro.
Revisores/as	Introducir al sistema de tweets catalogados según la escala de intensidad. Se establecerá un codebook para ello.
Usuarios/as	Visualizar el discurso de odio según filtros. Realizar análisis cualitativos. Proponer incorporación de palabras / aportar otras informaciones o comentarios
Técnico/a	Inicialmente para verificar que los procesos de captura y análisis son adecuados, este proceso se activará de forma manual para controlar que no hay errores en su ejecución y en su análisis. Una vez determinado que el correcto funcionamiento, se ejecutarán de forma automática en cadencias temporales.

Procesamiento de datos

Visualizar los estadísticos correspondientes al discurso de odio contenido en grandes volúmenes de tweets en tiempo real, puede comportar un tiempo de cálculo elevado, lo que implicaría que la aplicación funcionase muy lentamente. Para evitar esto, se pre-procesarán los resultados, de manera que se creará un fichero para cada día, con los resultados de los estadísticos e indicadores de los tweets correspondientes a ese día. La visualización del histórico de datos del discurso de odio, se hará a partir de estos ficheros preprocesados y no a partir de todos los tweets, lo que implicará una gran mejora en la eficiencia de cálculo y visualización de datos.
En cuanto al análisis cualitativo del discurso de odio, se propone poner las orientaciones para realizar dicho análisis, herramienta y orientaciones a disposición de los interesados, sin un procesamiento previo de los resultados generados.

Interfaz de usuario (visualización de datos)

Para poder visualizar los resultados de discurso de odio se diseñará una interfaz gráfica, para que el usuario pueda seleccionar y especificar los resultados deseados. Básicamente, los parámetros a seleccionar corresponderán a los parámetros de identificación diseñados en el proyecto y descritos en los apartados anteriores.
Durante la fase piloto se irán redefiniendo los requerimientos de visualización de los resultados del análisis de discurso de odio, así como el diseño de la interfaz gráfica que mejor se adapte a las necesidades de los diferentes usuarios de la aplicación. La aplicación puede permitir la visualización de:

1. Indicadores y estadísticos principales de los resultados del discurso de odio (según los filtros establecidos y que sean viables).
2. Tabla y gráfico de evolución de discurso de odio (número de tweets) según frecuencia temporal (día/semana/mes/año) y filtros seleccionados.
3. Tabla y gráfico de nube de palabras frecuentes según los filtros seleccionados.
4. Exportar la lista de tweets (.xls) según filtros seleccionados.
5. Elaboración de informes predefinidos de forma automática.

10

Referencias

PROTOCOLO
Y SISTEMA DE INDICADORES
para la detección del discurso de odio en las redes sociales

10. Referencias

Austin, J. (1961). Cómo Hacer cosas con palabras. Palabras y acciones (Paidós). Barcelona.

Burnap, P., & Williams, M. L. (2015). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeli ng for Policy and Decision Making. Policy and Intenet, 7(2), 223–242.

Burnap, P., & Williams, M. L. (2016). Us and them: identifying cyber hate on Twitter across multiple protected characteristics. EPJ Data Science, 5(1). https://doi.org/10.1140/epjds/s13688-016- 0072-6

Cabré-Olivé, J., Flecha-García, R., Ionescu, V., Pulido, C., & Sordé-Martí, T. (2017). Identifying the Relevance of Research Goals through Collecting Citizens' Voices on Social Media. International and Multidisciplinary Journal of Social Sciences, 6(1), 70. http://diposit.ub.edu/dspace/bitstream/2445/126740/1/678905.pdf

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering, 10(4), 215–230. https://preventviolentextremism.info/sites/default/files/A%20Lexicon-Based%20Approach%20for%20Hate%20Speech%20Detection.pdf

Greevy, E., & Smeaton, A. F. (2004). Classifying racist texts using a support vector machine (p. 2).

Association for Computing Machinery (ACM). https://doi.org/10.1145/1008992.1009074 Hughey, M. W., & Daniels, J. (2013).

Racistcommentsatonlinenewssites: Amethodologicaldilemmafordiscourseanalysis. Media, Cultureand Society,35(3), 332–347. https://journals.sagepub.com/doi/10.1177/0163443712472089

Miró Llinares, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. IDP. Revista de Internet, Derecho y Política, 22(22). https://dialnet.unirioja.es/servlet/articulo?codigo=5849356

Noriega, C. A., & Iribarren, F.J. (2012). Social Networks for Hate Speech: Commercial Talk Radio and New Media. UCLA Chicano Studies, 2. Retrieved from https://www.chicano.ucla.edu/publications/report-brief/social-networks-hate-speech

Observatorio Proxi. (2015). Informe del Observatorio Proxi. Barcelona. Retrieved from https://www.observatorioproxi.org/images/pdfs/INFORME-proxi-2015.pdf

Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. PLOS ONE, 13(8), e0203117. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0203117

Searle, J., & Soler, M. (2004). Lenguaje y ciencias sociales. Barcelona: El Roure.

Van Dijk, T. A. (2002). Discourse and racism. In D. T. Goldberg & J. Solomos (Eds.), A Companion to Racial and Ethnic Studies. Oxford: Blackwell.

Van Dijk, T. A. (2007). Discurso racista. In Igartua, J.; Muñiz, C. (2007) Medios de comunicación, Inmigración y Sociedad. (pp. 9–16). Salamanca: Universidad de Salamanca. Retrieved from http://eps-salud.com.ar/Pdfs/Van_Dijk_Discurso_Racist

Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. IEEE Access, 6, 13825–13835. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8292838

Wieviorka, M. (1992). El espacio del Racismo. Barcelona: Paidós.

ANEXO 1

Creación del banco de palabras

ANEXO 1

El análisis del discurso de odio utiliza en la mayoría de los casos, aunque no sea suficiente, listados de palabras que denoten o tiendan a denotar odio. Por este motivo, en el análisis de artículos y de las herramientas hemos tratado de identificar "bancos de palabras ("bag of words" en la literatura) que pudieran utilizarse en la herramienta de monitoreo y detección del discurso de odio. A continuación, se describe la metodología desarrollada para la creación de un banco de palabras específico para el proyecto.

1. Revisión de bancos de palabras existentes y lista de palabras unificada

Tanto en el análisis de artículos científicos como de las herramientas, se ha intentado obtener las "listas de palabras" utilizadas para la identificación del discurso de odio. Sin embargo, los bancos de palabras de las herramientas que hemos explorado y de artículos consultados, en su mayoría, no están disponibles, hecho que de acuerdo con Fortuna y Nunes (2018) ralentiza los avances en este campo. En algunos casos, los artículos o documentación relacionada con herramientas existentes describen los métodos para elaborar una bolsa de palabras, pero no se facilita la lista resultante. Muchos de los artículos analizados ponen ejemplos de palabras o expresiones pero no ofrecen una relación completa del vocabulario utilizado.

Así mismo, el idioma es una dificultad añadida, dado que nuestro objetivo es obtener o crear un banco de palabras en español/castellano, y en el caso de obtener palabras en inglés, no son traducibles directamente por el uso muy específico del lenguaje, a menudo propio del argot de cada idioma.

De los artículos revisados que han creado léxicos, hemos revisado cuáles están disponibles, como el caso de Mathew et.al. (2018) y cuáles no, como en Gitari et. al. (2015). Algunos estudios recientes ya parten de tweets y anotaciones manuales realizadas y disponibles por ejemplo en Crowdflower (**https://data.world/crowdflower/hate-speech-identification**) y el repositorio de código abierto Github (**https://github.com/ZeerakW/hatespeech**), que están también en inglés (los léxicos de estos recursos son prácticamente inaplicables en castellano).

En las herramientas en que estaba disponible, hemos recopilado los léxicos (bancos de palabras) y traducido las mismas al castellano. Concretamente, hemos obtenido las siguientes listas:

a. Banco de palabras a partir de Hate meter
b. Banco de palabras a partir de Hate base
c. Banco de palabras a partir de Hate meter (hashtags)
d. Banco de palabras a partir de PROXI

ANEXO 1

a. Banco de palabras a partir de Hate meter

Página Web: **http://hatemeter.eu**

Fecha de recogida: Abril 2019

Metodología de obtención: consulta del listado propuesto y se ha comprobado que se utiliza en castellano.

Palabras: Isis; Terrorismo; Fundamentalismo; Jihad; Bastardo; Guerra; Islam; Mohammed; Halal; Islam; Ramadan; Stop Invasión; Invasión; Invasores; Violencia; Sangre; Tolerancia cero; Sumisión; Guerra Civil; Ciudadanía; Integración; No integración; jóvenes musulmanes; Cultura; Migrantes; Expulsión; Acogida; Clandestinos; Africanos; Stop Inmigración; Recursos; Asilo; Refugiados; No Eurabia; Civilización; Patria; EU; Europa; Fronteras; Primero Nuestra Gente; Occidente; comunistas; No liberales; No democracia; Partidos; Idiotas; Mujeres; Izquierda; Burca; Niqab; Violación; Papa; ONG; Gitanos; Negros; Nigerianos; Ebreos; Judíos; Latinos; China; Filipinos; Bangladesh

b. Banco de palabras a partir de Hate Base

Página Web: https://hatebase.org/
Fecha de recogida: Abril 2019

Metodología de obtención: consulta del listado propuesta y se ha comprobado que se utiliza en castellano. Posteriormente, se solicitó a 3 personas de diferentes Comunidades Autónomas (Madrid, Andalucía, Cataluña) y una persona de una asociación musulmana que revisaran la lista obtenida para descartar las palabras que no conocían. Se han mantenido todas las palabras que han sido incluidas por lo menos por una de las personas participantes.

Palabras: amariconada; amariconadas; amariconado; **amariconados**; amariconar; **amariconé**; **amaricono**; **blanco**; **bollera**; **bolleras**; **cabezas cuadradas**; cabecita negra; **catalufo**; **catalufos**; **champinon; charnega**; **charnego**; **chele**; **chelo**; conguito; **conguitos**; **culandrón**; **culandrones**; **disminuido**; **disminuidos**

ANEXO 1

c. Banco de palabras – Hatemeter (hashtags)

Página Web: http://hatemeter.eu/
Fecha de recogida: Abril 2019

Metodología: se obtuvo una lista de hashtags en inglés relacionados con terrorismo, religión, invasión, identidad nacional/europea, múltiples grupos diana. Cada hashtag se ha traducido y probado en twitter para comprobar si se utiliza en lengua española. Se han mantenido en esta lista los hashtags que efectivamente son utilizados en castellano.

Terrorismo: #Isis; #Terrorismo (#Terrorism); #Terrorista (#Terrorists); #Fundamentalista (#Fundamentalists); #Jihadista (#Jihadists); #Jihad; #Seguridad (#Security); #AllahAkbar; #Atentado #AtaqueTerrorista; #Bastardo (#Bastards); #Guerra (#War); #HermanosMusulmanes (#Brotherhood, Una de las más importantes organizaciones islamistas internacionales); mención de ciudades/ Países donde han ocurrido atentados terroristas como: #Paris (#Paris), #Alemania (#Germany); #Münster; #Francia #France); #Belgica (#Belgium).

Religión: #Islam; #Sharia; #Coran (#Quran); #Mahoma (#Mohammed, a menudo asociado con #hashtags #Pedofilo – #Pedofilia, acusado de haberse Casado con una niña); #Halal; #Islam; #Mezquita (#Mosques); #AsiaBibi (Mujer cristiana, condenada por blasfemia en Paquistán en 2010); #Ramadan; #Halal. Frecuentemente contrapuestos a #Cristianos (#Christians); #Catolicos (#Catholics); #Cristiandad (#Christianity).

Invasión: #Stopinvasion (#Stopinvasion), #Invasion (#Invasion); #Invasores, (#Invaders); Violencia (#Violence); #Sangre (#Blood); #masacre (#Bloodbath); #ToleranciaCero (#ZeroTolerance); #Submisión (#Submission); #Guerracivil (#Civicwar).

Integración Social: #Ciudadania (#Citizenship); #Integración (#Integration); (#Sons, referring to the "second generation"); #NoIntegración (#NoIntegraton); #Jovenimusulman (#YoungMuslims); #Cultura (#Culture).

Inmigración: #Migrant (#Migrants); #Expulsión (#Expelled); #Acogida (#Hospitality); #Clandestino (#IllegalImigrants);#Africanos (#Africans); #Africa; #TodosenCasa; #TodosVueltalPaisdeOrigen #EverybodyGoBacktotheirCountriesofOrigin);

ANEXO 1

#StopInmigracion (#StopImmigration); #Recursos (#Resources, referido a los migrantes; es una expresion acuñada por el centro-derecho); #CerrarPuertos (#Closetheports); #DemandanteAsilo(#AsylumSeekers); #Aeropuertos (#Landings); #Refugiados (#Refugees).

Identidad National/ Europea: #DefensaEuropa; #NoEurabia; #Civilización (#Civilization); #España; #Españoles; #Patria (#Homeland); #EU; #Europa (#Europe); #Frontera (#Borders); #PatriotaEspañol (#ItalianPatriots); #PrimerolosdeCasa (#FirtsOurPeople); #Occidente (#West).

Multiple victims target group: centro-izquierda coalición (#Comunista – #Communists; #Buonismo – #BleedingHearts; #Izquierdabuonista– #BleedingHeartLiberals; #Buenismooccidental –#NOPD – #NODemocraticParty; Condiciones Mujer: #Mujer (#Women); #Burqa; #Burka; #Niqab; #Violación – #Rape; Non- governmental organisations (#ONG). Minorias sociales (#Rom – #Gitano; #Negro – #Niggers; #Ebreos – #Judios; #Latinos; #Bangla; #China; #Filipinos).

d. Banco de palabras a partir de PROXI

Página Web: http://www.observatorioproxi.org/
Fecha de recogida: Abril 2019

Metodología: Se ha recogido las nubes de palabras que tienen publicadas y que están relacionadas con el objetivo del proyecto.

Expresiones:
INMIGRANTES: “reparto de refugiados”, en lugar de acogida; “inmigrantes interceptados, detenidos y retenidos”, como si fueran delincuentes, “inmigrantes de segunda generación”, “extranjeros con doble nacionalidad”, refiriéndose a ciudadanos españoles, “ilegales”, “sin papeles”, “indocumentados”, e incluso “carga humana”, “avalancha”, “oleada” o “marea” de inmigrantes, “Aquí no hay trabajo para todos”, “Pues mételos en tu casa”, “Vienen a vivir de

ANEXO 1

las ayudas", "Son unos delincuentes, parásitos, mendigos", "Todos son del top manta, gorrillas, lateros…", El "Efecto llamada", "Los musulmanes son potenciales terroristas", "Los musulmanes no se quieren integrar", "Nos están invadiendo", "Aquí no hay trabajo para todos", "Mételos en tu casa",
GITANOS: "Son unos parásitos", "No se quieren integrar", "Atracan con droga", "Viven de las ayudas, abusando del sistema", "No respetan las normas de convivencia", "Son unos parásitos".

e. Banco de palabras a partir de Wikcionario – Categoría ES – Términos despectivos

Página Web: **https://es.wiktionary.org/wiki/Apéndice:1000_palabras_básicas_en_español**

Fecha de recogida: Abril 2019

Metodología: En el enlace se encuentra un listado de términos despectivos, que contiene términos racistas. Pero no está dividido por grupo al que va dirigido.

A partir del estado de la cuestión, se realizó una lista unificada de posibles palabras de odio, eliminando las duplicidades, y se ordenó alfabéticamente.

ANEXO 1

2. Grupo de discusión para la selección de palabras

Se realizó un grupo de discusión para seleccionar las palabras relevantes para el banco de palabras. El grupo de discusión estuvo formado por 10 Investigadores e investigadoras de diferentes edades y contextos geográficos españoles. Se discutieron todas las palabras de la lista y se crearon 2 categorías de palabras:

I. Palabras que por sí solas denotan odio.
II. Palabras que pueden ser utilizadas en discursos de odio, y que por lo tanto dependen del contexto o de la relación con otras palabras.

La selección de palabras y categorización en una u otra categoría se hizo por consenso respecto a cada término, entendiendo que este era un punto de partida.

ANEXO 1

3. Análisis cuantitativo de perfiles racistas e incorporación de palabras

Otra estrategia ha consistido en obtener un listado de palabras de un conjunto de perfiles de usuarios de twitter que eran de manera obvia y ostentosa racistas. Para ello:

Primero, se identificó un listado de 20 perfiles, a partir de hashtags o expresiones obviamente racistas, islamófobas o antisemitas. Sólo se seleccionaron perfiles españoles y escritos en español. Se excluyeron algunos perfiles que eran muy politizados. La propuesta fue contrastada por 3 personas del equipo. Los perfiles identificados son:

Cuenta perfil	Tipo de odio	Número tweets timeline	Número seguidores	Likes
@cuenta1	Racistas - Islamofobia	986	206	1327
@cuenta2	Racistas - Islamofobia	4777	152	1402
@cuenta3	Racistas - Islamofobia	43.2K	2646	50.1K
@cuenta4	Racistas - Islamofobia	2781	321	3856
@cuenta5	Racistas - Islamofobia	48K	3399	34.8K
@cuenta6	Racistas - Islamofobia	134K	2973	87.1K
@cuenta7	Racistas - Islamofobia	6700	128	10.6K
@cuenta8	Racistas - Islamofobia	21.9K	26.1K	25K
@cuenta9	Antisemita	14.7K	4452	26.8K
@cuenta10	Antisemita	4394	485	10.9K
@cuenta11	Antisemita	2609	199	2197

ANEXO 1

Cuenta perfil	Tipo de odio	Número tweets timeline	Número seguidores	Likes
@cuenta12	Antisemita	2673	507	364
@cuenta13	Antisemita	7281	4004	44.7K
@cuenta14	Antisemita	1196	35	287
@cuenta15	Antisemita	105K	3775	27.3K
@cuenta16	Antisemita	5537	126	2626
@cuenta17	Antisemita	11.5K	5411	1880
@cuenta18	Antisemita	16.4K	7171	1758

Se programó una aplicación con el lenguaje de programación Python para la captura de tweets de estos perfiles y su análisis. La captura se ha realizado a través de las API's de Twitter: (**https://developer.twitter.com/en/docs**). Se capturaron todos los tweets de los perfiles seleccionados y se analizó su contenido en dos sentidos: Por una parte, se ha realizado un análisis de frecuencias de hashtags en todo el conjunto de los tweets de los perfiles seleccionados, que ha permitido crear un listado de los hashtags más utilizados. Por otra parte, se ha realizado un análisis de frecuencias de las palabras de los mensajes de los tweets. Para este análisis se han utilizado librerías y conjuntos de palabras de parada estándares para español, y se han rechazado manualmente palabras correspondientes a símbolos o abreviaciones. Los hashtags y palabras identificados.

Hashtags frecuentes en los usuarios racistas		
bds: 841	defensemddhh: 121	paro: 59
stopinvasion: 497	barcelona: 107	atletijuegolimpio: 57
israel: 461	últimahora: 93	stopbulos: 54

ANEXO 1

Hashtags frecuentes en los usuarios racistas		
palestina: 350	eleccionesya: 92	corrupcion: 54
boycotteurovision2019: 319	artistasqueyadijeronno: 89	freepalestine: 49
stopislam: 306	ddhh: 87	pleasedontgo: 49
closeborders: 288	spexit: 81	nakba: 45
elecciones2019votoderechas: 272	openarms: 76	womentogaza: 45
gaza: 210	maccabiesapartheid: 72	matisyahu: 45
apartheid: 181	ivreich: 68	aquarius: 43
vox: 176	forolocalbds: 68	templarios: 41
openborders: 169	españa: 67	politicalcorrectness: 39
ue: 166	venezuela: 66	defiendeespaña: 37
españaviva: 150	merkel: 65	concertperpalestina: 37
eurabia: 135	nakba70: 64	ot18galafinal: 37
yonocomproapartheid: 130	palestine: 64	elai: 37
ongs: 126	bdsesddhh: 60	

ANEXO 1

Palabras frecuentes en los usuarios racistas				
españa: 2028	tener: 391	venezuela: 262	sionistas: 201	votar: 165
ser: 1704	mal: 391	nacional: 262	ningún: 201	@dioshorus796: 165
israel: 1389	todas: 389	violencia: 260	persona: 200	religión: 164
: 1266	tal: 388	sistema: 259	hora: 200	grande: 164
@vox_es: 1139	@bdspaisvalencia: 387	andalucía: 259	mundial: 200	ó: 164
: 1132	alguien: 385	@orbitaeduardo: 255	madre: 200	des: 164
gracias: 1123	@jguaido: 385	haber: 254	@larecolectiva: 200	liberal: 164
años: 1104	https: 381	frente: 254	veces: 199	buen: 163
@marubimo: 1077	quiere: 381	musulmanes: 253	podría: 199	ataque: 163
así: 1069	tras: 379	@elhadadevox: 253	cristianos: 198	ue: 163
ahora: 1021	año: 378	tres: 252	ilegales: 198	: 163
vox: 1011	historia: 377	dar: 252	mierda: 197	régimen: 162
hoy: 1004	sino: 377	cosa: 251	quiero: 196	subvenciones: 162
the: 978	según: 376	derecho: 250	@ortega_smith: 196	mira: 162
va: 950	dicen: 370	cierto: 249	@anebald: 195	manera: 162
bien: 918	presidente: 370	cualquier: 249	@mimariban: 195	actual: 162
españoles: 915	dios: 369	siendo: 248	@el_pais: 194	ciudad: 162
hace: 911	barcelona: 366	for: 247	saber: 194	horas: 161
per: 859	usted: 366	ilegal: 247	artículo: 194	meses: 161
#bds: 847	millones: 365	gaza: 246	@elnahu_atr: 194	partidos: 161

ANEXO 1

Palabras frecuentes en los usuarios racistas				
@rubnpulido: 834	psoe: 362	veo: 246	tema: 193	pronto: 161
ver: 788	quieren: 361	forma: 244	queda: 193	solidaritat: 161
puede: 779	els: 361	mayoría: 242	@irrintzialaves: 193	poner: 160
día: 775	@bdsmadrid: 358	trump: 242	general: 192	: 160
gente: 774	inmigración: 358	boicot: 242	seguro: 191	@sandonaegui: 160
hacer: 749	cosas: 357	solidaridad: 238	foto: 190	normal: 159
país: 731	: 356	policía: 237	control: 189	calle: 159
@santi_abascal: 722	política: 352	espero: 236	dia: 189	onu: 159
aquí: 715	judío: 352	@mariagtriana: 236	alguna: 189	libro: 159
solo: 709	#palestina: 352	medio: 235	terroristas: 189	ong: 159
dice: 707	apoyo: 350	vergüenza: 234	deja: 189	demonios: 159
palestina: 707	m: 346	casi: 233	@mnopasana: 189	acuerdo: 158
amb: 707	niños: 345	sabe: 232	saben: 188	última: 158
vez: 706	caso: 344	pasado: 229	demás: 188	mes: 158
sólo: 706	bueno: 342	paz: 228	humanos: 187	puedes: 158
mismo: 702	nunca: 341	semana: 228	buena: 187	encima: 157
vía: 694	és: 341	on: 227	gustó: 187	nombre: 157
gobierno: 689	libertad: 339	dijo: 227	peor: 186	puesto: 157
inmigrantes: 654	sociedad: 338	dels: 227	imagen: 186	siguen: 156
verdad: 638	casa: 331	justicia: 226	habla: 185	bandera: 155
cada: 637	ejemplo: 330	defensa: 226	paso: 185	voto: 155

ANEXO 1

Palabras frecuentes en los usuarios racistas				
europa: 634	is: 329	canal: 226	político: 185	poble: 155
personas: 625	igual: 328	francia: 225	libre: 184	noticia: 154
menos: 622	sigue: 328	familia: 225	pide: 183	vosotros: 154
pues: 613	países: 327	soros: 225	@tiradorfranco13: 183	razón: 154
pueblo: 604	izquierda: 325	haciendo: 224	#apartheid: 182	rey: 154
dos: 603	vamos: 325	seguridad: 224	tierra: 182	sale: 154
vídeo: 601	internacional: 324	población: 223	cambio: 182	fuerza: 154
@idealismonazi: 596	#boycotteurovision2019: 322	iglesias: 223	primero: 181	entidad: 153
siempre: 595	: 321	sionismo: 222	mano: 181	ejército: 153
parte: 595	quién: 320	primera: 222	judía: 181	fronteras: 153
gran: 591	pp: 318	bajo: 222	feminismo: 181	@carrascomarimar:153
mundo: 581	debe: 317	@e_cycni: 222	palestinos: 181	través: 152
mujer: 570	campaña: 310	ciudadanos: 221	defender: 181	are: 151
mejor: 567	pedro: 309	@psoe: 221	total: 181	idea: 151
@bdscatalunya: 558	favor: 308	programa: 221	islam: 181	género: 151
decir: 543	#stopislam: 306	twitter: 220	avui: 181	situación: 151
da: 535	hacen: 304	visto: 220	palabras: 180	lucha: 151
cómo: 533	ir: 304	: 220	datos: 180	posible: 150
podemos: 532	ley: 302	derecha: 219	movimiento: 180	ninguna: 150
van: 524	medios: 302	entonces: 219	israelià: 179	amigos: 150

ANEXO 1

Palabras frecuentes en los usuarios racistas				
mujeres: 516	@jadouken10: 302	alemania: 219	lleva: 178	centro: 150
vida: 502	además: 300	final: 217	#vox: 177	primer: 150
mientras: 500	nueva: 298	ello: 217	ayuda: 177	#españaviva: 150
#stopinvasion: 497	hombres: 298	saludos: 216	@elmundoes: 177	sabes: 149
@sanchezcastejon:493	pasa: 297	muerte: 215	padre: 177	htt: 149
sánchez: 493	@salvameoficial: 296	hijos: 214	cultura: 176	euros: 148
parece: 492	hablar: 292	lugar: 214	unas: 175	not: 148
judíos: 491	cataluña: 291	@hermanntertsch: 214	suport: 175	you: 148
@rescop1: 482	aunque: 290	l'apartheid: 214	algún: 174	militar: 148
hombre: 478	madrid: 289	incluso: 213	existe: 174	vídeos: 148
guerra: 477	elecciones: 288	nación: 213	@cama1610: 174	respecto: 148
claro: 474	ayer: 288	@edmondd09082129: 213	with: 173	liberales: 148
mas: 473	#closeborders: 288	social: 212	dan: 173	dejar: 148
tan: 473	ahí: 287	@casoaislado_es: 212	marruecos: 173	acto: 148
después: 470	fin: 286	#gaza: 212	miedo: 173	llegar: 148
#israel: 462	hilo: 285	momento: 211	@cristinasegui_: 173	@historiaespanna: 148
israelí: 452	mañana: 284	com: 211	palestí: 173	: 148
español: 451	vaya: 283	palestino: 210	invasión: 172	drets: 148
tiempo: 451	muchas: 282	hacia: 210	nivel: 171	discurso: 147
partido: 450	: 282	real: 209	maduro: 171	blanco: 147

ANEXO 1

Palabras frecuentes en los usuarios racistas				
hecho: 442	problema: 277	dentro: 209	gusta: 170	señor: 147
luego: 436	falta: 277	grupo: 209	viene: 169	@varyingweion: 147
toda: 436	dicho: 275	único: 207	público: 169	sociales: 146
nadie: 435	ve: 275	http: 206	:/: 169	varios: 146
@youtube: 432	seguir: 274	tipo: 206	#openborders: 169	luz: 146
apartheid: 432	misma: 273	: 206	grandes: 168	feminista: 146
dinero: 422	odio: 271	voy: 205	eeuu: 168	refugiados: 146
sido: 419	#elecciones2019votoderechas: 271	ddhh: 205	información: 168	sr: 146
cuenta: 418	pueden: 270	lado: 204	pedir: 168	chile: 146
sionista: 418	trabajo: 269	hola: 204	@danaeon_: 168	@solof1sincirco: 146
poder: 413	@voxnoticias_es: 269	més: 204	origen: 167	feministas: 145
@rodrickgamer: 413	derechos: 268	digo: 203	tampoco: 167	zona: 145
realidad: 407	políticos: 268	hijo: 203	jajaja: 167	iglesia: 145
video: 405	cara: 265	debería: 203	@ino_forever: 167	aquest: 145
creo: 403	aún: 265	@a3noticias: 203	comunidad: 166	número: 144
@agnosis9: 399	@proucomplicitat: 265	civil: 202	plan: 166	cabeza: 144
días: 398	mayor: 264	mismos: 202	bds: 166	amigo: 144
nuevo: 395	sé: 264	pablo: 202	#ue: 166	dado: 143
española: 393	etc: 263	llama: 201	democracia: 165	puedo: 143

ANEXO 1

4. Análisis cualitativo de tweets racistas e incorporación de palabras

Finalmente, el equipo de investigación ha realizado análisis cualitativos de tweets que contienen discurso de odio, para identificar otras palabras frecuentes a incorporar en el listado inicial. Se han revisado más de 1500 tweets obtenidos con la estrategia de captura top-down por colectivos descrita más arriba.

Las estrategias descritas han dado lugar a nuestro banco de palabras. Este banco tiene 2 categorías de palabras para identificar discurso de odio (Palabras que por si solas denotan odio; Palabras que pueden ser utilizadas en discursos de odio).

ANEXO 2

Ejemplos de tweets

ANEXO 2

Tweets con palabras propias del Banco de palabras

Presencia de palabras propias del Banco de palabras	Sudaca vuelvete al cono sur, eres un puto charnego
	Por cierto, los gilipollas son los progres de Podemos, PSOE y demás basura y tooooooooda la banda maricomplejines de PP y C's, pues sus "líderes" están encantados con esto de la invasión programada y la Eurabia que se nos viene encima.
	Detenidos dos machupichus por estafar 28.000 euros a un discpacitado

ANEXO 2

Tweets clasificados según su intensidad en discurso de odio

Odio extremo	Discurso que incita a la violencia.	Hay que matarlos a todos jatejode, un negro de mierda mas un negro menos, no se pierde nada.
		que negro re de mierda wacho. Después los hijos de mil putas se quejan cuando dicen de matarlos a todos a estos hijos de puta
		Que asco de gente!! A estos hijos de perra habría que colgarlos de los huevos hasta que mueran. Así de claro. Como hacen ellos. Ni expulsión ni hostias. Pena de muerte.
		Palestina resiste! Viva el pueblo palestino, Viva su heroica resistencia! Muerte a los sionistas!
		COTO UNA AYUDITA PARA ELIMINAR A LOS JUDIOS. UN ABRAZO Y GRACIAS
Odio - ofensa	Discurso que representa ofensas personales o colectivas, que incita a la discriminación, reproduce tópicos y falsedades.	Hay que joderse, otro extranjero que quiere que España se vaya a la mierda, que asco de gentuza, seguro que algo les debe a estos golpistas catalufos
		Que dices del "Pueblo Elegido" invasor y asesino del pueblo Palestino? Que dices de Angela Merkel que se ha reelegido en 3 ocasiones y va por la 4a? Que dices de los Jeques árabes que nadie eligió y gobiernan desde hace décadas...
		Gracias al #RamadanMubarak los délitos de #menas se concentrarán fuera del horario de ayuno
Discurso neutro	Discurso descriptivo, en el que no aparece odio.	¿Sabes cómo diagnosticar el #eccema en personas de piel negra? ¿ #SabíasQue las personas negras son más propensas a desarrollar formas más severas de eczema que las personas de otras etnias? #Dermatología https://www.medicalnewstoday.com/articles/325066.php …
Upstander	Discurso alternativo, que contribuye a una contra-narrativa, rompiendo con los tópicos o posicionándose en defensa de los colectivos objeto de odio.	Yo soy gitano y no soy delincuente
		¿Cómo te sonaría "el clan mafioso de argentinos", "el clan mafioso de judíos", "el clan mafioso de porteños"? Estigmatizar por religión, identidad, procedencia social o territorial se llama Xenofobia. Por cierto, parece que t... https://twitter.com/gabicerru/status/1127206684655529985, https://twitter.com/gabicerru/status/1127206684655529985
		Bulo sobre actos delictivos de un grupo de menores en Calella de Mar. Hemos contactado por teléfono con Policía Local de Calella de Mar y Mossos d'Esquadra de Pineda de Mar y confirman que ha habido hurtos pero el resto de la información es FALSA.

ANEXO 2

Tweets clasificados según el tipo de lenguaje

Lenguaje que incita o realiza apología de la violencia	Yo diferencio legalidad de JUSTICIA, estoy absorto de la capacidad de aceptación de la población, por ejemplo cuando un moro viola a una hija o sobrina. Yo sin duda haría justicia y luego entraría en Soto, pero feliz. No se si me explico Lynn
	La hermana de un amigo sale de rendir de recibirse y viene un negro y le roba el celular... Ah pero uno dice que a los negros hay que matarlos apenas nacen y te miran mal.
	No seais fachas. Es solo un pobre inmigrante que ha dejado su pais de origen y su familia en busca del sueño español. MECAGOENSUPUTAMADRE. Lo que le hace falta a este mierda es una buena ensalada de ostias, que por lo que se ve, todavia no se las han dado.Y despues a su puto pais.
Justificación, bromas, trivialización de la violencia hacia 'los otros	El problema reside en que si ahora uno le parte la cara al moro, saldría en todos los telediarios con el titular de ataque racista de un blanco a un pobre y desvalido musulman, recordemos lo que pasó en Barcelona con el vigilante de seguridad del tren,que fue hasta sancionado.
	Si todo el mundo que dice ser superviviente del Holocausto lo es de verdad, entonces a quien mato Hitler? **#hitler #holocuento**
	Ojalá el Karma tape la boca a ésa sinvergüenza, ósea q está diciendo q x ser extranjero y negro tienen derecho a matar,violar y apalear pués para q su familia o ella sea la próxima en recibir tan maravilloso trato.
Lenguaje divisorio o de otredad (Ellos versus Nosotros)	Seguro k era un moro??que casualidad si es que Barcelona ya casi parece Casablanca.
	Yo os traduzco: los menas tendrán en Cataluña una paga de 665 euros mensuales hasta los 23 años. El efecto llamada será brutal. Cuando os roben, cuando os atraquen, cuando os agredan o violen, volved la cabeza hacia el Parlamento de Cataluña y en las próximas elecciones haced algo.
	Por culpa de los negros que violan y matan mujeres nos tratan a todos los hombres como el enemigo. Pues NO, el enemigo está en el gobierno y es el que deja impune estos actos y permite que cualquier inmigrante se quede en nuestro país y pueda hacer lo que le de la gana.

ANEXO 2

Reproducción prejuicios de estereotipos	Peleas tribales, comidos y pagados por los españoles, se dedican más cómodamente a lo que han estado haciendo durante más de 2000 años a peleas tribales por el control.….Claro que de un puntapié los problemas va a parar al otro lado de la verja..y todos tranquilos #Stopinvasion [Link a noticia sobre pelea entre marroquís en Ceuta]
	Por esos gitanos que van a la autoescuela en su propio coche!!
	Hemos votado que queremos seguir trayendo moros para que sigan haciendo esto y viviendo de nuestras ayudas. Es lo que hay, ahora toca joderse.
Difusión de rumores	Este año desgraciadamente ha habido 20 manadas, solo se habla de esta, las otras víctimas no tienen derechos pues han sido violadas por moros sudacas y demás escoria.
	Nooooo, así no es la noticia, la noticia es QUE OTRO EXTRANJERO, MORO, la más señas, a si el causante de esta muerte, y ya van unas cuantas muertes, violaciones y maltrato a mujeres por parte de esta gentuza que vosotros seguís protegiendo y ayudando,sois una banda SINVERGÜENZAS.
	Y siguen viniendo este finde vinieron tres cientos como no cierren las fronteras vamos de culo y a la ruina se necesita control ayudas a la gusrdia civil para proteger las fronteras.
Difusión de datos falsos (False Facts)	Este año desgraciadamente ha habido 20 manadas, solo se habla de esta, las otras víctimas no tienen derechos pues han sido violadas por moros sudacas y demás escoria.
	En Móstoles una manada de moros, para variar, trataron de violar a una chica y al salir sus amigos en su defensa estos animales les atacaron con navajas. No habrá ni manifestaciones feministas ni a ministra dirá de cambiar el código penal. Más asco no podéis dar.
	Sin querer hacerme el experto, Europa Occidental está más cerca de Europistán que otra cosa. Lo que no se dice en los medios es terrible. Alemania tiene colegios donde el alumnado nativo ya no resiste porque es atormentado por los alumnos musulmanes.

ANEXO 2

Argumentos Trampa (Flawed argumentation)	La sociedad española es una de las más seguras. El asesino de Parla es un moro magrebí con antecedentes, uno de ellos por tentativa de homicidio, que habría sido deportado si gobernara #VOX. La gran mayoría de los violadores y asesinos son de origen inmigrante. Sois cómplices. pic.twitter.com/6Y2cmMxD74
	Acabo de oir en la radio, que la pelea entre marroquíes y sudamericanos se debió a que los putos sarracenos intentaron violar a una chica dominicana. Está claro que el problema, no es la inmigración, el problema son los moros y su cultura
	Esa etnia por casualidad no será la misma que lleva decenios, sino siglos, saltan2las leyes de los payos, robando, engañando y mintiendo a los payos, pero al mismo tiempo queriendo dar pena para vivir del cuento. Hay señol guardia sivil que me han roba do la fregoneta. Haaay ¡
Metáforas-comentarios deshumanizadoras-ironías	Menas alojados en hotel de Calella, se espera un plácido verano allí
	No es lo mismo que un judío se bañe a que me bañe con un judío
	Ya mismo soy un musulmán me tocas y exploto culiaaaa
Estructuras no-verbales despectivas. Ej. Emoticonos, puntuación, mayúsculas, etc.	JARED LETO PARECE UN MUSULMAN LPM EN CUALQUIER MOMENTO GRITA ALLAHU AKBAR, DE CARA PARECE JESUS Y NI HABLAR DE QUE ANDA A SABER XQ LLEVO UNA CABEZA FALSA #MetGala #MetGala2019 pic.twitter.com/UTUaTyh6RH
	Continúa en tu HOLOCUENTO, asno repugnante JUDIO, descendiente de SIÓN. Sigue viendo y creyendo en la “la vida es bella” o “el niño con la pijama de rayas”. Ponte a leer el verdadero holocausto: El holocausto PALESTINO, por parte de los de tu especie. “Fiesta del Purim”

ANEXO 2

Tweets según diferentes modalidades de lenguaje Upstander

Publicación de comentarios propios: argumentación con evidencias desmintiendo prejuicios Y visibilización de experiencias positivas.	Yo soy gitano y no soy delincuente
Intervenciones antirumor: datos y explicaciones contra-rumor.	Bulo sobre actos delictivos de un grupo de menores en Calella de Mar. Hemos contactado por teléfono con Policía Local de Calella de Mar y Mossos d'Esquadra de Pineda de Mar y confirman que ha habido hurtos pero el resto de la información es FALSA.
Intervenciones pedagógicas: hechos y datos para clarificar debate.	Más de 68'5 millones de personas viven fuera de sus hogares por la guerra, la violencia y graves violaciones de sus derechos fundamentales. Esto supone el número más alto jamás registrado **#DíaMundialDelRefugiado** **#RefugeeDay**
Intervenciones sensibilización: apelar a las emociones.	¿Cómo te sonaría "el clan mafioso de argentinos", "el clan mafioso de judíos", "el clan mafioso de porteños"? Estigmatizar por religión, identidad, procedencia social o territorial se llama Xenofobia. Por cierto, parece que t… **https://twitter.com/PatoBullrich/status/1126865474015330304**
Cualificación de comentarios de otros usuarios: valorar positivamente o descalificar comentarios positivos/ negativos.	"Dos clases de menas". Qué penita joder. Qué penita. Hay dos clases de adultos: quienes creen en los chavales y les dan apoyo bajo las circunstancias que sean y quienes no. Os cuento una historia en este hilo
Denuncia de comentarios de odio: a través de las herramientas propias de las redes sociales y apps creadas específicamente para ello.	Todavía quedan otras elecciones. Queremos cubrirlas y estar preparad@s para los bulos que van a intentar colarnos, como que "el peligro es el inmigrante". ¿Nos ayudas a seguir? **https://twitter.com/ctxt_es/status/1125777557465419776 pic.twitter.com/SIfVQJx4vz**

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