

Índice

- — Redes sociales y archivos sociales
Redes sociales de uso general: presente y futuro
- — Fases metodológicas:
 - PLANTEAMIENTO INVESTIGADOR
 - RECUPERACIÓN DE INFORMACIÓN DESDE REDES SOCIALES
 - CONCEPTUALIZACIÓN, METODOLOGÍAS DE ANÁLISIS
 - EXPLOTACIÓN Y DISEMINACIÓN DE RESULTADOS
- — Un ejemplo de investigación:
Redes sociales & Humanidades Digitales en #BLM
- — Conclusiones y debate

Muchos procesos sociales y culturales tienen su reflejo y expresión en las **redes sociales**.

Las redes sociales como termómetro social, objeto de estudio de procesos humanísticos (.e.g patrimonialización, memoria, identidad, conexión pasado-presente, derechos humanos & amenazas a democracia)

Los usuarios expresan opiniones, debaten, comparten contenido y planifican acciones, como manifestaciones o procesos participativos.

El controvertido “archivo social”

What is critical is not the medium by which both objects are archived, but the act of archiving itself: that is, that moment of ‘forgetting’ that foregrounds the set of historical, material and ideological set of practices that makes a collection possible.

(Pybus 2013) regarding Bishop and O’Driscoll (2004)

¿ES UNA FUENTE, UN MEDIO, UN SOPORTE?
¿QUÉ ARCHIVAMOS: DATOS, PROCESOS...?

¿QUÉ VALOR TIENEN?
¿QUÉ GUARDAN/QUÉ REFLEJAN?

User profiles do not ‘collect the dust of statements that have become inert’ but are rather continuously (re)filled within more and more information, allowing users to be simultaneously involved and exploited by something that is larger than themselves.

From this perspective, the archive is not simply of mnemonic or historical importance but has a biopolitical function that allows for the production or social, political and economic subjectivities, which actively circulate within social networks. (Pybus 2013)

El controvertido “archivo social”

FONDOS, ARCHIVEROS Y POLÍTICAS (Y LA PROBLEMÁTICA DEL CONSENTIMIENTO)

(...) includes ‘fonds’ but also ‘archivists’ (...) and the ‘policies’ defined for the arrangement of records and ‘archives’. Hence, the term ‘social archive’ does not refer to a ‘social media’ collection, but to an archive built by social and not just ‘professional’ consent.

Social fonds: the entire body of records created by a community of users through a participatory dynamic on social platforms around a shared interest or event. (Ruiz Gómez and Vallés 2020)

DEFINICIÓN DE ARCHIVO SOCIAL NO CONSENSUADA RETOS METODOLÓGICOS IMPLICACIONES SOCIALES (Y PRÁCTICAS DENTRO DE NUESTRA INVESTIGACIÓN)

Algunos archivos sociales HD

- Aspectos conductuales, de memoria y olvido
(Youtube as archive Gehl, 2009; Acker, 2014)

Gehl, R. (2009). YouTube as archive: Who will curate this digital Wunderkammer? *International Journal of Cultural Studies*, 12(1), 43-60. <https://doi.org/10.1177/1367877908098854>

- Estudiar aspectos de violencia de género
(Vallès and Soler Jiménez, 2016; Ruiz Gómez and Vallès, 2020)

#Cuéntalo: the path between archival activism and the social archive(s) Vicenç Ruiz Gómez and Aniol Maria Vallès



Algunos archivos sociales HD

- Patrimonio y Racismo
(Otero and Martín-Rodilla, 2021)

- Contextos de lenguaje y enfermedad mental
(Parapar, Martín-Rodilla, Losada and Crestani, 2021).

RESEARCH-ARTICLE

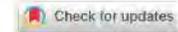


Building Cultural Heritage Reference Collections from Social Media through Pooling Strategies: The Case of 2020's Tensions Over Race and Heritage

Authors: [David Otero](#), [Patricia Martín-Rodilla](#), [Javier Parapar](#) [Authors Info & Claims](#)

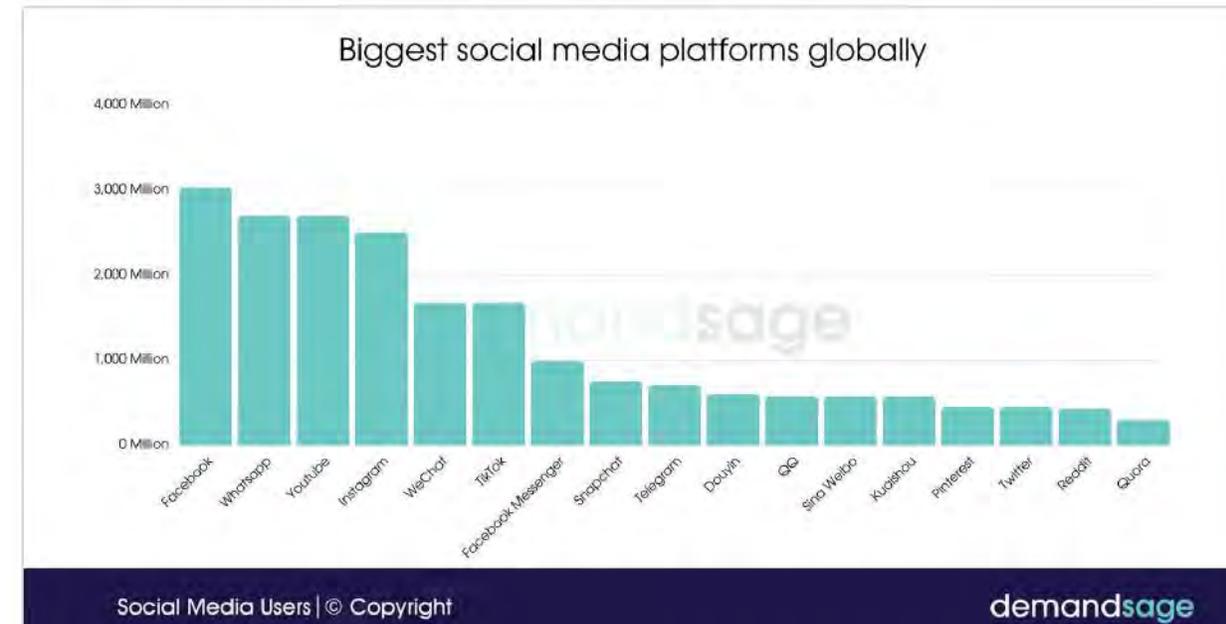
Journal on Computing and Cultural Heritage, Volume 15, Issue T • Article No.: 9, pp 1–13 • <https://doi.org/10.1145/3477604>

Published: 14 December 2021 [Publication History](#)



Redes sociales de uso general: presente y futuro

This table shows the biggest social media platforms globally as of 2023:



- 4.900 millones de usuarios de redes sociales: el 60,49% de la población mundial.
- Se prevé que los usuarios de redes sociales alcancen los 5.850 millones en 2027.
- La tasa de penetración global de las redes sociales es del 59,4%.
- Un usuario típico de redes sociales interactúa con 6,6 plataformas de redes sociales.
- Las redes sociales son utilizadas por el 85% de los 5.270 millones de usuarios de teléfonos móviles del mundo.
- Una persona usa las redes sociales durante dos horas y treinta y cinco minutos todos los días de media.
- Millennials y Generación Z son los usuarios más frecuentes de las plataformas de redes sociales.
- Tendencia: TikTok (aumento de usuarios del 105% en Estados Unidos durante los últimos dos años.)

Fuente: <https://www.demandsage.com/social-media-users/>

Redes sociales de uso general: presente y futuro



Cada app de terceros conectada a la API de Instagram tiene restricciones de acceso. Recientemente se ha cambiado este acceso de 5000 a 200 peticiones por hora.

<https://eclincher.com/instagrams-api-limitations-2022>



Todas las solicitudes a la API están sujetas a límites de frecuencia. Las solicitudes a la API Graph y la [API de visualización básica de Instagram](#) están sujetas a los [límites de frecuencia de la plataforma](#), mientras que las solicitudes a la [API de marketing](#) y la [API Graph de Instagram](#) están sujetas a los [límites de frecuencia de casos de uso comercial \(BUC\)](#).

Current X limits

The current technical limits for accounts are:

- **Direct Messages (daily):** The limit is 500 messages sent per day.
- **posts:** 2,400 per day. The daily update limit is further broken down into smaller limits for semi-hourly intervals. Reposts are counted as posts.
- **Changes to account email:** 4 per hour.
- **Following (account-based):** Once an account is following 5,000 other accounts, additional follow attempts are limited by account-specific ratios.

These limits include actions from all devices, including web, mobile, phone, API, etc. API requests from all third-party applications are tracked against the hourly API limit. People who use multiple third-party applications with their account will therefore reach the API limit more quickly.

These limits may be temporarily reduced during periods of heavy site usage. In such cases, we will post an update on the [X status](#) site.



Accesos gratuitos a la API:

- 100 consultas por minuto (OAuth client id si usas OAuth)
- 10 consultas por minuto si no usas OAuth

La mayoría de las apps de terceros caen en estos rangos y pueden usar la API sin coste.

<https://www.redditinc.com/policies/data-api-terms>, Junio 2023



Find the right access for you

Free

For write-only use cases and testing the X API

- Rate limited access to v2 post posting and media upload endpoints
- 1,500 Posts per month - posting limit at the app level
- 1 app ID
- Login with X
- Free

[Get started](#)

Basic

For hobbyists or prototypes

- Rate limited access to suite of v2 endpoints
- 3,000 Posts per month - posting limit at the user level
- 50,000 Posts per month - posting limit at the app level
- 10,000 Posts per month - read-limit rate cap
- 2 app IDs
- Login with X
- \$100 per month

[Subscribe now](#)

Pro

For startups scaling their business

- Rate-limited access to suite of v2 endpoints, including search and filtered stream
- 1,000,000 Posts per month - GET at the app level
- 300,000 Posts per month - posting limit at the app level
- 3 app IDs
- Login with X
- \$5,000 per month

[Subscribe now](#)

Enterprise

For businesses and scaled commercial projects

- Commercial-level access that meets your and your customer's specific needs
- Managed services by a dedicated account team
- Complete streams: replay, engagement metrics, backfill, and more features
- Monthly subscription tiers

[Apply now](#)



Y las reacciones...

Study warns API restrictions by social media platforms threaten research

Research on people and society to suffer as social media platforms limit access to user data

Press release

Published on Thursday 2 November 2023

Last updated on Thursday 2 November 2023

[View more announcements in Communications](#)



Media enquiries

Alison Jones
aj250@bath.ac.uk

Media Centre

[Find information for journalists, bloggers and online commentators](#)

Comment | [Published: 02 November 2023](#)

Platform-controlled social media APIs threaten open science

[Brittany I. Davidson](#) , [Darja Wischerath](#), [Daniel Racek](#), [Douglas A. Parry](#), [Emily Godwin](#), [Joanne Hinds](#), [Dirk van der Linden](#), [Jonathan F. Roscoe](#), [Laura Ayravainen](#) & [Alicia G. Cork](#)

[Nature Human Behaviour](#) (2023) | [Cite this article](#)

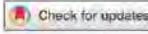
376 Accesses | 86 Altmetric | [Metrics](#)

Social media data enable insights into human behaviour. Researchers can access these data via platform-provided application programming interfaces (APIs), but these come with restrictive usage terms that mean studies cannot be reproduced or replicated. Platform-owned APIs hinder access, transparency and scientific knowledge.

Submit an article

Journal homepage

401

COMMUNICATION METHODS AND MEASURES
<https://doi.org/10.1080/19312458.2023.2181319> Routledge
Taylor & Francis GroupOPEN ACCESS 

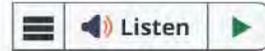
Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking

Jakob Ohme^{a,b}, Theo Araujo^b, Laura Boeschoten^{b,c}, Deen Freelon^d, Nilam Ram^e, Byron B. Reeves^e, and Thomas N. Robinson^f

^aWeizenbaum Institute for the Networked Society, Freie Universität Berlin, Berlin, Germany; ^bAmsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands; ^cDepartment of Methodology and Statistics, Utrecht University, Utrecht, Netherlands; ^dHussman School of Journalism and Media, University of North Carolina at Chapel Hill, Raleigh, North Carolina, USA; ^eDepartment of Communication, Stanford University, Stanford, USA; ^fDepartments of Pediatrics and Medicine, Stanford University, Stanford, USA

ABSTRACT

In social media effects research, the role of specific social media content is understudied, in part attributable to the fact that communication science previously lacked methods to access social media content directly. Digital trace data (DTD) can shed light on textual and audio-visual content of social media use and enable the analysis of content usage on a granular individual level that has been previously unavailable. However, because digital trace data are not specifically designed for research purposes, collection and analysis present several uncertainties. This article is a collaborative effort by scholars to provide an overview of how three methods of digital trace data collection - APIs, data donations, and tracking - can be used in studying the effects of social media content in three important topic areas of communication research: *misinformation*, *algorithmic bias*, and *well-being*. We address the question of how to collect raw social media content data and arrive at meaningful measures with multiple state-of-the-art data collection techniques that can be used to study the effects of social media use on different levels of detail. We conclude with a discussion of best practices for the implementation of each technique, and a comparison of their advantages and disadvantages.

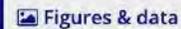
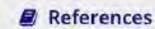
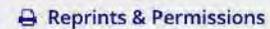
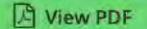
 Listen

Forum

The Data Abyss: How Lack of Data Access Leaves Research and Society in the Dark

Claes de Vreese  & Rebekah Tromble

Pages 356-360 | Published online: 20 May 2023

 Cite this article  <https://doi.org/10.1080/10584609.2023.2207488> Full Article Figures & data References Citations Metrics Reprints & Permissions View PDF View E

Intentando responder a...



¿Qué estrategias hemos explorado para trabajar con redes sociales & humanidades digitales? ¿Cuáles no?



¿Qué retos plantea introducir esta fuente en una investigación en HD?

Estrategias en 4 Fases:

PLANTEAMIENTO INVESTIGADOR

RECUPERACIÓN DE INFORMACIÓN
DESDE REDES SOCIALES

CONCEPTUALIZACIÓN,
METODOLOGÍAS DE ANÁLISIS

EXPLOTACIÓN Y DISEMINACIÓN DE
RESULTADOS



PLANTEAMIENTO INVESTIGADOR

- Preguntas de investigación y objetivos siempre en el contexto humanístico. Pensar enlaces presente-pasado (¿Qué intentamos ver ahora y como conecta con antes? Viceversa?)
- Sistemáticamente abordar:
 - Fuentes posibles y fuentes existentes: ¿Tiene sentido extraer (más) de redes sociales? ¿Qué aporta? ¿Podemos re-usar?
 - Idiomas, procedencia.
 - Sesgos de la fuente: consultar y contextualizar con estudios sobre tipos de usuarios. ¿Quiénes están (y no) en (cada) rrs?
- Documentar todo esto si generamos nuevas colecciones.



RECUPERACIÓN DE INFORMACIÓN DESDE REDES SOCIALES

¿CÓMO GENERALMENTE SE RECOGE, ARCHIVA Y PROCESA ESTA INFORMACIÓN?

¿QUÉ TÉCNICAS Y PROTOCOLOS SE USAN EN SU AUTOMATIZACIÓN?

Técnicas de RI manuales o semi-manuales

Técnicas de RI automáticas no adaptadas a dominios sociales y humanísticos

POOLING = Método para extraer una muestra de un conjunto de documentos (posts o publicaciones en el caso de redes sociales), obteniendo el conjunto más completo e imparcial de documentos relevantes sobre un tema determinado.



RECUPERACIÓN DE INFORMACIÓN DESDE REDES SOCIALES

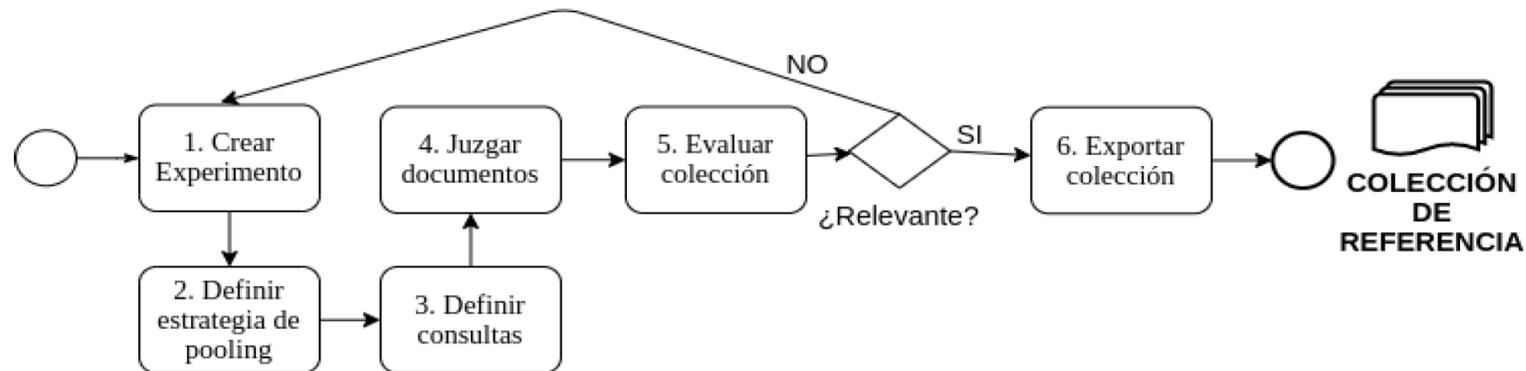




RECUPERACIÓN DE INFORMACIÓN DESDE REDES SOCIALES

NUESTRA PROPUESTA:

- Metodología de aplicación de pooling como estrategia RI desde redes sociales para la construcción de colecciones de referencia en contextos humanísticos y sociales.
- Caso de estudio: colección de referencia desde *Reddit* para el estudio del impacto en la comunidad hispanohablante en redes sociales del movimiento Black Lives Matter y las revueltas a raíz del asesinato de George Floyd en 2020.





RECUPERACIÓN DE INFORMACIÓN DESDE REDES SOCIALES

- NUEVAS LIMITACIONES Y COSTES EN RECUPERACIÓN
- FORMACIÓN INVESTIGADORA HÍBRIDA
- ARCHIVOS SOCIALES PARCIALES Y DISEÑADOS PARA UN PROPÓSITO/PREGUNTAS DE INVESTIGACIÓN CONCRETAS

RETOS:

- EXPANSIÓN SEMÁNTICA, DEPENDENCIA LINGÜÍSTICA EN LA CONSTRUCCIÓN
- DATOS PERSONALES Y PROCESOS DE ANONIMIZACIÓN; CONSENTIMIENTO Y COMITÉS ÉTICOS
- REUSADO FUERA DE LA INVESTIGACIÓN VIGENTE AÚN POR EXPLORAR
- INTEROPERABILIDAD CON OTROS ARCHIVOS SOCIALES, REPOSITORIOS O DATASETS CON OBJETIVOS SIMILARES O COMPLEMENTARIOS



CONCEPTUALIZACIÓN Y METODOLOGÍAS DE ANÁLISIS

- Conceptualización explícita y en consonancia con las preguntas de investigación y la estrategia de recuperación (Muy difícil “volver atrás”)
- Análisis en varios ejes: cualitativo vs. cuantitativo, outliners vs. predominancia
- Gran volumen de datos en análisis: no improvisar!
- La importancia de la Visualización
- ¿Qué podemos automatizar o qué procesos son adecuados para someter a mi colección para resolver (al menos inicialmente) mis preguntas?

CONCEPTUALIZACIÓN Y METODOLOGÍAS DE ANÁLISIS



RETOS:

- CONCEPTUALIZACIÓN ¿COMPARTIDA?
- SELECCIÓN DE TÉCNICAS DE ANÁLISIS:
 - NLP: niveles léxico, gramáticas, discursivo/argumental. Análisis de sentimiento, etc. (Dependiente del idioma)
 - GRAFOS
 - CONTENIDO MULTIMEDIA
 - COMPORTAMIENTO E INTERACCIÓN, ETC.
- ¿LAS RRSS SON UNA FUENTE MÁS? JUSTIFICACIÓN Y PUESTA EN CONTEXTO DEL ARCHIVO SOCIAL Y SU ANÁLISIS POSTERIOR.

EXPLOTACIÓN Y DISEMINACIÓN DE RESULTADOS



- LIBERACIÓN DE ARCHIVOS Y COLECCIONES

Resources & Links

- METACOLECCIONES



Contact

Patricia Martín-Rodilla, PhD.

Assistant professor & Information Retrieval Lab (IRLab) researcher

EXPLOTACIÓN Y DISEMINACIÓN DE RESULTADOS



RETOS:

- COSTES “DE MANTENIMIENTO”
- CALIDAD
- REUSADO: ¿DIFÍCIL?, ¿CON CONDICIONES?

- PRODUCTO NO “PUBLICABLE”

**BLACK
LIVES
MATTER**

PLANTEAMIENTO INVESTIGADOR: RESUMEN

Caso de Estudio: archivo social durante las tensiones raciales de 2020

- Asesinato de George Floyd mayo de 2020. Condena junio de 2021.
- Oleadas de protestas primero en EEUU, después Inglaterra y finalmente Europa y Sudamérica.
- Movimiento #BLM organizado desde redes sociales: convocatoria de concentraciones y ATAQUES A ENTIDADES PATRIMONIALES consideradas símbolos racistas.
- La conexión entre racismo y manifestaciones patrimoniales ha sido ampliamente estudiada y revisada por el grupo en artículos previos*
- *Reddit* como termómetro social del proceso

PLANTEAMIENTO INVESTIGADOR

- Objetivos
- Preguntas de investigación

Sistemáticamente abordar:

- Fuentes posibles y fuentes existentes: ¿Tiene sentido extraer (más) de redes sociales? ¿Qué aporta? ¿Podemos re-usar?
- Idiomas, procedencia.
- Sesgos de la fuente: consultar y contextualizar con estudios sobre tipos de usuarios. ¿Quiénes están (y no) en (cada) rrss?

- ¿Qué conexiones hay entre el movimiento actual #BLM y las diferentes manifestaciones patrimoniales atacadas? ¿A qué procesos de memoria responden y de qué colectivos/personas/lugares?
- ¿Qué eventos pasados tienen vigencia en la memoria actual y se explicitan en los ataques?
- ¿Hay cambios en las percepciones de patrimonio a lo largo del tiempo en este escenario?

RECUPERACIÓN DE INFORMACIÓN

Caso de Estudio: archivo social durante las tensiones raciales de 2020

1. Definición de términos de búsqueda (proceso paralelo en inglés y con pruebas de cobertura). El pooling permite la redefinición y replicación de esta fase.
2. Juicios de relevancia por expertos: **219** posts juzgados por **dos anotadores** (1 hora/evaluador). Estudio de acuerdo inter-anotación en marcha.
3. Descarga del contenido *Reddit* (entre 25 de mayo de 2020 y 10 de abril del 2021) que contenían publicaciones anotadas positivas: 77 hilos diferentes, con un total de **19303 posts en español** en la colección de referencia.

racismo	vandalizado
protesta	vandalismo
racista	esclavo
George Floyd	esclavitud
blm	policía
Black lives matter	violencia
estatua	monumento
busto	revuelta
placa	

RECUPERACIÓN DE INFORMACIÓN



- Casi 1 año de actividad en Reddit de la comunidad hispanohablante acerca de las protestas antirracismo y su expresión en entidades patrimoniales.
- Media de 250 posts incorporados a la colección por hilo anotado.
- Información de *subreddits* de toda Latinoamérica, España y comunidades hispanas en otros países, con al menos 15 países diferentes identificados. Cobertura en *subreddits* no directamente relacionados con la temática.

CONCEPTUALIZACIÓN Y METODOLOGÍAS DE ANÁLISIS

➤ **Subpreguntas: ¿Qué entidades y personajes se atacaron? ¿Por qué? ¿En qué localizaciones? ¿Fueron cambiando con el tiempo? ¿Qué perfiles de usuario encontramos (quién y cómo habla sobre #BLM)? ¿Qué eventos pasados podrían relacionarse en las localizaciones?**

- **NLP: niveles léxico, gramáticas, discursivo/argumental. Análisis de sentimiento, etc. (Dependiente del idioma)**
- **SELECCIÓN DE TÉCNICAS DE ANÁLISIS**
 - **GRAFOS**
 - **CONTENIDO MULTIMEDIA**
 - **COMPORTAMIENTO E INTERACCIÓN, ETC.**
- **¿LAS RRSS SON UNA FUENTE MÁS? JUSTIFICACIÓN Y PUESTA EN CONTEXTO DEL ARCHIVO SOCIAL Y SU ANÁLISIS POSTERIOR. Ampliación a otras redes sociales + estudio de representatividad y sesgo. Cualquier fenómeno social es más de lo que se ve/se escribe/se refleja/se organiza en RRSS.**

EXPLOTACIÓN Y DISEMINACIÓN DE RESULTADOS

¿SABEMOS (AHORA) ALGO MÁS SOBRE EL #BLM QUE ANTES?

- Localizaciones con tradición esclavista y/o tensión racial en diversos eventos previos generan más ataques/interacciones en Reddit. Grandes clústeres identificados
- La edad y el género como factor
- Próximo paso: análisis de árboles RST de discurso

¿COMO ENCAJA ESTO CON OTRAS FUENTES? ¿SE CONFIRMA?

EXPLOTACIÓN Y DISEMINACIÓN DE RESULTADOS

COLECCIONES #BLM Anonimizadas disponibles en abierto

Resources & Links



<https://pages.irlab.org/david.otero.freijeiro/hdh2021/>

Contact

Patricia Martín-Rodilla, PhD.

Assistant professor & Information Retrieval Lab (IRLab) researcher

@Martín-Rodilla, P. Curso de verano 2024 USC

Lenguaje en contextos de enfermedad mental

- Preguntas de investigación y objetivos siempre en el contexto humanístico. Pensar enlaces presente-pasado (¿Qué intentamos ver ahora y como conecta con antes? Viceversa?)
 - Objetivos
- Ejemplos:
- Detección precoz estudiando patrones a todos los niveles (léxico, gramatical, discursivo, etc.)
 - Diferencias entre patologías, comorbilidad y lenguaje
 - Perfiles automáticos con mayor riesgo
 - Diferencias multilingües
 - Evolución temporal

Sistemáticamente abordar:

- Fuentes posibles y fuentes existentes: ¿Tiene sentido extraer (más) de redes sociales? ¿Qué aporta? ¿Podemos re-usar?
- Idiomas, procedencia.
- Sesgos de la fuente: consultar y contextualizar con estudios sobre tipos de usuarios. ¿Quiénes están (y no) en (cada) rrss?



En resumen...



Hemos explorado estrategias en 4 fases
**(planteamiento investigador, IR,
conceptualización & análisis y
explotación y diseminación)**



Hemos identificado (algunos) retos que
plantea introducir RRSS como fuente
en una investigación en HD

**Con este marco de trabajo, las RRSS pueden
constituir una fuente valiosa en HD**

Explorando las redes sociales como fuente de investigación en HD

aproximaciones para enlazar el presente al pasado

Patricia Martín-Rodilla

Científica Titular, Humanidades Digitales. CSIC

p.m.rodilla@iegps.csic.es

**MOITAS
GRAZAS!**