



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Escuela de Ingeniería Informática de Oviedo

Proyecto fin de Máster

Comportamiento problemático en medios sociales: ¿es el Fediverso la solución?

Ángel García Menéndez

Tutor:

Daniel Gayo Avello

Noviembre de 2023

Resumen

Este proyecto ahonda en la evolución de los medios sociales, desde sus orígenes descentralizados en plataformas como BBS, UseNet o IRC hasta el surgimiento de las redes sociales cerradas y privativas, conocidas como “jardines vallados”. En el proyecto se exponen las críticas alrededor de los medios sociales privativos, incluida la mercantilización de los datos de los usuarios y su impacto en la salud mental y bienestar de estos. Se explora la aparición de alternativas, especialmente el Fediverso, con una atención especial a Mastodon como alternativa viable a Twitter. Se analizan las interacciones con periodistas presentes en ambas plataformas, pues son uno de los principales objetivos del acoso en línea. El proyecto presenta un análisis de profundidad, comparando el discurso de los usuarios en ambas plataformas. Los resultados muestran diferencias en la riqueza y uso del vocabulario, aportando información sobre las diferentes experiencias de usuario en estas dos redes sociales. Se concluye haciendo hincapié en la necesidad de más indagación para poder discernir si esas diferencias implican un cambio positivo o negativo y en la búsqueda de metodologías alternativas para la recolección de datos.

Palabras clave: medios sociales, Twitter, Mastodon, discurso en línea, acoso, análisis de contenidos

Abstract

This paper delves into the evolution of social media, from its decentralized origins in platforms like BBS, UseNet, and IRC to the rise of privately owned and closed social media networks, known as “walled gardens”. The paper discusses the criticisms surrounding closed social media, including the commodification of user data and its impact on user well-being and mental health. The emergence of alternatives, particularly the Fediverse, is explored, with a focus on Mastodon as a viable alternative to Twitter. Interactions with journalists present in both platforms are analyzed, as they are one of the main targets of online harassment. The paper presents an in-depth analysis comparing user discourse in the two sites. The results reveal differences in vocabulary richness and usage, providing insights into the distinct user experiences in these two social media environments. The paper concludes by emphasizing the need for further research to assess the positive or negative nature of these differences and explore alternative data-gathering methods in privacy-conscious environments.

Keywords: social media, Twitter, Mastodon, online discourse, harassment, content analysis

A Papá y a Teresa, por permitirme llegar aquí.

A Ana, por dar luz a los días oscuros.

A Dani, por ser un gran tutor aunque él piense lo contrario.

A Patru y Migue, por su ayuda con la literatura psicológica.

*A los compañeros del Máster, que cedieron sus máquinas para poder
descargar el dataset.*

Índice general

1	Introducción	4
1.1	Motivación	4
1.2	Finalidad	6
2	Objetivos	7
3	Estado del arte	9
4	Metodología	11
4.1	<i>Dataset</i>	11
4.2	Preprocesamiento	12
4.3	Análisis	14
5	Resultados	17
6	Conclusiones y trabajo a futuro	22
7	Dirección y gestión del proyecto	24
7.1	Planificación del proyecto	24
7.1.1	Identificación de interesados	24
7.1.2	OBS y PBS	24
7.1.3	Planificación inicial (EDT)	28
7.1.4	Gestión de riesgos	36
7.1.5	Presupuesto inicial	39
7.2	Ejecución del proyecto	43
7.2.1	Plan de seguimiento de la planificación	43
7.2.2	Bitácora de incidencias	43
7.2.3	Seguimiento de riesgos	45
7.3	Cierre del proyecto	50
7.3.1	Planificación final	50
7.3.2	Informe final de riesgos	50
7.3.3	Presupuesto final	50
7.3.4	Informe de lecciones aprendidas	58

ÍNDICE GENERAL

Apéndices	63
A Plan de gestión de riesgos	64
A.1 Metodología	64
A.2 Roles y responsabilidades	64
A.3 Presupuesto	64
A.4 Calendario	65
A.5 Categorías de riesgo	65
A.6 Definiciones de probabilidad	66
A.7 Definiciones de impacto por objetivos	66
A.8 Matriz de probabilidad de impacto	66
A.9 Tolerancias	66
A.10 Formatos de la documentación	67
A.11 Seguimiento	67
B Cuestiones éticas	68

Capítulo 1

Introducción

Este proyecto nace como continuación de mi propio TFG [GM21]. En él, se hacía una retrospectiva de los medios sociales descentralizados y federados, basados en las tecnologías de software libre, estándares abiertos y la no pertenencia a una única organización o individuo.

En la sección 1.1 se vuelven a exponer algunos de los puntos de aquel trabajo y la razón por la que se afronta el actual. Después, en la sección 1.2 hablaremos de cuál es la principal contribución del artículo.

1.1 Motivación

Podría decirse que los medios sociales son anteriores incluso al propio concepto de “medios sociales”. BBS, UseNet o IRC, tecnologías que preceden a la web, podrían clasificarse como medios sociales [BM19], e incluso comparten algunos de sus conceptos clave con plataformas actuales. No obstante, a diferencia de las plataformas de hoy día, todas estas tecnologías compartían una característica común: ser descentralizadas. No estaban controladas por una única organización.

Sin embargo, a comienzos de los años 2000 una nueva ola de plataformas de medios sociales privados y propietarios hicieron su aparición. Entre ellos encontramos a MySpace en 2003, Facebook en 2004 y Twitter en 2006. Estos sitios ofrecían a sus usuarios una experiencia similar a las de las plataformas de las que antes hablábamos, pero ahora estaban controlados por compañías privadas buscando un beneficio económico.

Este fue el nacimiento de lo que el World Wide Web Consortium (W3C) bautizó como *walled gardens* o jardines vallados [Ye23]. Estos son servicios deliberadamente diseñados para mantener la atención de sus usuarios restringida a una única plataforma y, al mismo tiempo, aislarlos del resto de la Web. La principal

CAPÍTULO 1. INTRODUCCIÓN

preocupación del W3C viene del hecho de que estos jardines vallados constituyen una amenaza a los principios en los que se basa el ecosistema de Internet. Esto es, la interoperabilidad de tecnologías y estándares libres y abiertos.

A lo largo de los años han ido apareciendo diferentes críticas a este modelo de medios sociales. Algunas son principalmente ideológicas, como las de [De05], que sostiene que estas plataformas despojan de significado a la comunicación humana, la cual es transformada en un nuevo tipo de mercancía (datos de los usuarios) a explotar y monetizar. Ese es también el principal argumento de [Zu19], que estudia la vigilancia a la que son expuestos los usuarios de estos sitios y cómo la incesante búsqueda de mayor interacción entre los usuarios pasa factura a esos mismos usuarios. Esto, por otro lado, parece estar teniendo efectos negativos en la salud mental de los usuarios, tal como se ha explorado en [AJNP20], [BLM22], [BB17] e incluso en investigaciones internas por parte de Instagram [WHS21].

Como respuesta a esto, se han realizado diversos intentos de generar alternativas a los medios sociales cerrados, como podría ser el proyecto Open Social de Google y MySpace. No obstante, el punto de inflexión llegó con el lanzamiento de Identi.ca, más tarde rebautizado a StatusNet, y la subsiguiente implementación del estándar OStatus [W3]. Este protocolo del W3C allanó el camino para la interoperabilidad entre los sitios web de medios sociales, permitiendo a los usuarios interactuar entre ellos desde diferentes plataformas de forma transparente.

El último paso fue la creación de ActivityPub, un estándar maduro del W3C para la implementación de plataformas de medios sociales interoperables [WT18]. Es alrededor de esta tecnología cómo nace el Fediverso (Universo Federado), una colección de plataformas sociales interconectadas mediante el protocolo ActivityPub. De entre todas ellas habría que resaltar Mastodon, un sitio de *microblogging* similar a Twitter.

Aunque el proyecto de Mastodon comenzó su andadura en 2016, su popularidad como alternativa a Twitter se disparó cuando esta última fue comprada por el magnate Elon Musk en octubre de 2022. De acuerdo a estadísticas de Google, la popularidad de la consulta “mastodon” aumentó un 97 % ese mes. Es más, de acuerdo a algunas estimaciones, Mastodon podría haber alcanzado los 8 millones de usuarios a finales de noviembre de 2022¹. Ha habido incluso investigaciones académicas sobre el fenómeno de la “migración de Twitter”, y por qué Mastodon parece ser una de las alternativas más escogidas [Va23].

¹Fuente: mastodon.social/@mastodonusercount. Es necesario puntualizar que esta cifra es discutida, pues depende enormemente de la forma en la que se decida contar los usuarios de una red descentralizada.

1.2 Finalidad

Dado el reciente desarrollo de los acontecimientos, con dos plataformas, Twitter y Mastodon, con características similares y un sustancial trasvase de usuarios de una a otra, tenemos ante nosotros una oportunidad real de comparar las dinámicas y comportamiento de los usuarios en cada plataforma, siendo cada una representativa de modelos diferentes de medios sociales.

Por tanto, atendiendo a las oportunidades que nos brindan estas circunstancias, la principal contribución del presente trabajo sería **dilucidar si Mastodon ofrece un entorno más favorable para los usuarios de medios sociales que Twitter**. En la sección 2 se plantea el cómo investigar esta cuestión.

Capítulo 2

Objetivos

Ya se ha establecido que el principal objetivo del trabajo es comparar la experiencia de los usuarios en Twitter y Mastodon para comprobar si el cambio de paradigma de este último ofrece una experiencia de medio social más sana y positiva, sin los impactos negativos que se exponían en la sección 1.1. Para ello se plantean 3 preguntas de investigación (RQ):

- **RQ1** ¿Hay profanidades presentes de forma similar en las interacciones de ambas plataformas?
- **RQ2** ¿Muestran las profanidades en las interacciones de ambas plataformas unos patrones de uso similares?
- **RQ3** ¿Comparten los usuarios de ambas plataformas un vocabulario similar al discutir sobre temas polémicos?

Aclarar que en este trabajo entendemos como profanidades el mal vocabulario en sentido amplio, incluyendo insultos, vejaciones, obscenidades, barrabasadas, &c. Los detalles de cómo dar respuesta a estas RQ se encuentran en la sección 4.

En lo que serían requisitos de proyecto, se establecen los siguientes:

1. Se estudiarán las interacciones con periodistas
2. Se descargarán las interacciones de las plataformas Twitter y Mastodon
3. Las interacciones estudiadas serán sólo aquellas en lengua inglesa
4. Se seguirán las recomendaciones de las guías éticas de la Association of Internet Researchers
5. Se empleará el lenguaje de programación Python
6. Se empleará la API académica de Twitter

CAPÍTULO 2. OBJETIVOS

7. Los resultados serán enviados a una revista dentro del índice JCR

Capítulo 3

Estado del arte

A la hora de examinar qué se ha investigado antes, cabe plantearse tres cuestiones principales. La primera sería cómo pueden compararse dos grupos de usuarios en medios sociales. La segunda correspondería a cómo escoger a qué usuarios comparar. Finalmente, la tercera giraría en torno a cuáles han sido los principales trabajos que han estudiado el ecosistema de Mastodon recientemente.

Existe literatura específica que explora la comparación de grupos de usuarios en diferentes medios sociales, como [LQ13], dónde se analiza cómo el discurso varía entre los usuarios de Facebook y Twitter. Empero, uno de los principales campos de aplicación de estas técnicas es el de la política.

En [SP15] se estudiaron las diferencias en la actividad en Twitter de usuarios Demócratas y Republicanos en EEUU, tratando de discernir potenciales disparidades psicológicas entre ambos grupos. En cierta medida, las “cámaras de eco” resultaron ser una de las razones detrás de esas diferencias¹.

A medida que el panorama político estadounidense se ha ido polarizando en los últimos años, el discurso político en la web ha seguido la misma línea. Esto se puede ver en [Gu21], dónde se analizan las diferencias en el lenguaje empleado por políticos de los dos principales partidos de EEUU durante el confinamiento de 2020. Un análisis similar, aunque más dilatado en el tiempo, lo encontramos en [BP22], dónde se ilustra la evolución en el discurso de políticos republicanos y demócratas.

En esto contexto de polarización en redes, uno de los grupos más afectados es el de los periodistas, especialmente a partir de 2020. Estudios como [Wa20], [Wa22], [Ho23] y [Uw23] nos permiten afirmar que el gremio de los periodistas

¹Las cámaras de eco también son el principal tema de [Ga18], junto al rol que los diferentes tipos de usuarios juegan en su creación.

es frecuentemente objeto de acoso y comportamiento “troll” por parte de cierto segmento de los internautas.

Dado el creciente grado de descontento con los medios sociales privativos, ha habido un crecimiento notable en el interés por las alternativas a estos, como es el Fediverso. Por ejemplo, [Ak22] comparó los diferentes métodos para examinar el comportamiento de los usuarios dentro de Mastodon, incluyendo el uso de *hashtags*, bots y el compartir publicaciones. No obstante, ese interés tuvo un especial repunte a partir de la adquisición de Twitter por parte de Elon Musk. [LCAT23] estudió el fenómeno de la “Migración de Twitter” y los factores que la han condicionado. De igual forma, [Je23] investigó, no sólo el proceso de la migración en sí, sino también los patrones de comportamiento de los usuarios tanto en Twitter como en Mastodon, estando estos últimos más centrados en la construcción de comunidades.

Capítulo 4

Metodología

Como hemos visto en la sección 3, los periodistas son uno de los principales objetivos del acoso en medios sociales. Por ende, nos pueden servir como un sistema de alerta temprana para poder medir el nivel de hostilidad de una plataforma social, pues ellos serán de los primeros en recibir hostigamiento, de la misma forma que los canarios sirven para detectar de forma temprana el gas tóxico en las minas de carbón.

Por tanto, para poder responder a las RQ planteadas en la sección 2, hemos recopilado un *dataset* con las interacciones con periodistas presentes tanto en Twitter como en Mastodon. A continuación se explica cómo se ha descargado el *dataset*, cómo se han normalizado los textos del mismo y los experimentos concretos llevados a cabo para responder a las RQ.

4.1 *Dataset*

Los periodistas seleccionados para el *dataset* son los que constan una lista pública de Reddit¹, donde figuran sus nombres de usuario en Twitter y Mastodon. Estos periodistas, al darse de alta en esta lista, expresaban su voluntad de ser descubiertos por la comunidad, de ser encontrados por los usuarios. Por tanto, se les consideró candidatos ideales para el estudio.

Para la descarga de las interacciones en Twitter se empleó la API Académica de la plataforma² para descargar las respuestas y los retweets con citas a los periodistas. En el caso de Mastodon el proceso de descarga de las respuestas

¹Fuente: https://www.reddit.com/r/Mastodon/comments/yq1yan/list_of_hundreds_of_journalists_to_follow_on

²La API Académica, en el momento en el que se usó, permitía la descarga de hasta 10 millones de tweets, otorgando además acceso al archivo completo de tweets desde 2006.

de los usuarios a los periodistas³ resultó más complejo. Para empezar, Mastodon tiene una arquitectura federada, similar a la del correo electrónico, donde los usuarios pueden registrarse en diferentes servidores (aquí llamados “instancias”) e interactuar entre ellos de forma casi transparente. Esto implica que las peticiones tuvieron que hacerse al servidor concreto dónde se alojaba cada toot (nombre que reciben las publicaciones en Mastodon) de cada periodista, pues sería el que alojaría una copia de todas las respuestas. Además, la API de Mastodon sólo permite recuperar un toot por petición, haciendo que se alcanzasen constantes límites de peticiones. Este problema se solventó empleando varias máquinas con diferentes direcciones IP, cada una de las cuales, a su vez, tenía varios hilos haciendo peticiones a instancias diferentes, tratando de paralelizar la descarga lo máximo posible.

El *dataset* cubre el período comprendido entre el 3 de octubre de 2022, momento en el que Elon Musk anuncia su compra de Twitter con un vídeo entrando con un fregadero en las oficinas de la compañía, hasta el 28 de Marzo de 2023, cuando comenzó el proceso de descarga.

Para respetar las expectativas de privacidad de la comunidad de Mastodon, en cada instancia de la que se recuperó al menos una publicación se realizó un meticuloso análisis de sus códigos de conducta y términos de uso. Si cualquiera de los dos documentos contenía cláusulas prohibiendo el acceso a los datos de sus usuarios para fines de minería de datos, raspado web o propósitos similares, los toots recuperados de dicha instancia se excluían del *dataset*.

El *dataset* final contiene 3.822.815 tweets y 14.805.462 toots. En adelante usaremos estos nombres para referirnos a los textos de cada plataforma en concreto o “publicaciones” para referirnos a textos con independencia de su origen.

4.2 Preprocesamiento

Todo los textos del *dataset* se normalizaron para eliminar todo el ruido posible, siguiendo la metodología propuesta por [GJ17]. Para ello, se empleó la biblioteca Spacy de Python. Se puede ver un diagrama del proceso completo en la figura 4.1 y un ejemplo de normalización de un texto ficticio en el cuadro 4.1.

Se reemplazaron los emojis y emoticonos por sus descripciones textuales. En el caso de los emojis se hizo uso de una *pipeline* especializada implementada en el entorno de Spacy, mientras que para los emoticonos se elaboró manualmente un diccionario que asociaba a cada emoticono su descripción correspondiente en la lista oficial de emoticonos de la Wikipedia.

Se realizó también enlazamiento de entidades con nombre (*Named Entity Linking* en inglés, en adelante NEL) de forma semiautomática. Este proceso incluyó

³Actualmente no existe un equivalente a la funcionalidad de citar en Mastodon y por tanto no hay textos de ese tipo en el *dataset* de esa plataforma.

Cuadro 4.1: Ejemplo de la normalización de un texto de ejemplo

Paso	Texto
Original	@journalist have u checked this statement from the gop toooo? here is the link https://example.com 😊 #politics
Eliminación de URL	@journalist have u checked this statement from the gop toooo? here is the link 😊 #politics
Eliminación hashtags	@journalist have u checked this statement from the gop toooo? here is the link 😊
Eliminación de usuarios	have u checked this statement from the gop toooo? here is the link 😊
Eliminación de emojis & emoticonos	have u checked this statement from the gop toooo? here is the link shy beaming face with smiley eyes
Corrección de faltas	have you checked this statement from the gop toooo? here is the link shy beaming face with smiley eyes
Corrección de letras excesivas	have you checked this statement from the gop too? here is the link shy beaming face with smiley eyes
NEL	have you checked this statement from the republican party too? here is the link shy beaming face with smiley eyes
Eliminación de palabras vacías	checked statement republican Party link shy beaming face smiley eyes
Lematización	check statement republican party link shy beam face smiley eyes

la detección de entidades con nombre, acrónimos, sintagmas nominales⁴ y los términos principales de acuerdo a TextRank⁵ de cada publicación. Después, filtramos ese diccionario con la lista de palabras vacías de Taurus⁶ para eliminar los términos no esenciales. También se eliminaron todos los términos con una frecuencia de aparición menor a 1200, pues por debajo de ese umbral se estimó que los términos eran carentes de sentido tras un análisis manual. Tras esto, se realizaron peticiones contra la API de Wikidata para identificar las entidades originales de cada término y nombre principal. Finalmente, la lista se comprobó manualmente para corregir errores. Terminado el proceso, los términos identificados como entidades en los textos se sustituyeron con su “designación canónica”. Por ejemplo, términos como *gop* y *republicans* se cambiaron

⁴Frases cuyo núcleo es un nombre, como “el mayor conglomerado tecnológico del mundo”.

⁵El algoritmo TextRank es una implementación de PageRank para el procesamiento del lenguaje natural, que detecta los términos principales de un determinado texto.

⁶Fuente: <http://snowball.tartarus.org/algorithms/english/stop.txt>

a republican party.

4.3 Análisis

Para analizar el *dataset* los textos fueron codificados en vectores de encaje léxico (en inglés *word embeddings*), como se propone en [Mi13], usando para ello la funcionalidad `word2vec` de la biblioteca Gensim de Python. Los análisis para responder a RQ1 y RQ2 se inspiraron en las investigaciones de [Wa14] sobre el mal vocabulario en medios sociales. Sin embargo, dado que estos autores no han hecho pública su lista de profanidades, se ha empleado una lista alternativa⁷. Para abordar RQ3 se ha recurrido a una colección de palabras polémicas asociadas con el comportamiento *troll* en Internet, como se propone en [SBVB22].

En el caso de la RQ1, se crearon dos modelos `word2vec` diferentes, uno para el *dataset* de cada red social. Hecho esto, comprobamos las profanidades que había en cada *dataset*, identificando aquellas presentes en un *dataset* pero no en el otro. Por ejemplo, *boner*, *chink* o *sissy* son términos que se encuentran exclusivamente en los textos de Twitter, mientras que *enby* sólo está presente en los de Mastodon.

Para afrontar las otra dos RQ fue necesario alinear los dos modelos para poder compararlos. Al haberse creado cada modelo por separado, las posiciones geométricas de los vectores en el espacio vectorial no coinciden, y por ende es necesario alinearlos. Para ello se empleó el Alineamiento de Procrustes⁸, como se sugiere en [HLJ16b].

Una vez alineados los modelos, la métrica escogida para la comparación fue la similitud coseno [Si01], aprovechando la representación vectorial de nuestro corpus. En nuestro caso concreto, se trata del producto escalar de dos vectores normalizados, cuyo resultado estará comprendido entre 0 (menos similar) y 1 (más similar). De esta forma se puede medir hasta qué punto dos palabras se están usando de la misma forma.

Afrontamos RQ2 calculando la similitud coseno entre las profanidades que estaban presentes en ambos modelos. La idea es ser capaces de dilucidar si ese vocabulario común, esas profanidades que se usan en ambas plataformas, se usan de la misma forma. Un claro ejemplo son los insultos racistas, que se pueden usar como ataque a personas racializadas o de forma irónica por parte de esas mismas comunidades.

Para RQ3 se empleó el enfoque propuesto en [HLJ16a]. En lugar de hacer una comparación entre los términos directamente, lo que se estudió fue el contexto

⁷Lista de profanidades: <https://github.com/dsojevic/profanity-list>

⁸Una técnica estadística que permite transformar, rotar y escalar uniformemente dos objetos (matrices en nuestro caso) se puedan comparar directamente en un espacio determinado, empleando para todo ello álgebra lineal.

de las palabras. Se calcularon los 10 términos más cercanos a cada palabra tanto en Twitter como en Mastodon, de acuerdo a la similitud coseno, para después quedarse sólo con los *vecinos* que coincidieran en ambas plataformas para una misma palabra. Tras esto, creamos un vector de similitud de segundo orden, cuyos elementos son la similitud coseno entre los términos comunes. El paso final fue calcular la distancia coseno de los vectores de similitud de segundo orden, lo que nos da una puntuación de 0 a 1 para cada término polémico. De esta forma, podemos comparar si la forma de abordar estos temas controvertidos es la forma en un plataforma y en otra.

En RQ1 y RQ2 lo que se examinaba era el uso de las profanidades y los cambios en su uso. En RQ3 no son los términos en sí lo que nos interesa, sino las discusiones que hacen aflorar. Al ser estos términos la causa de debates airados, queríamos comprobar la similitud de toda la conversación en torno a ellos, para ver si el debate se desenvuelve de la misma forma en Twitter y en Mastodon.

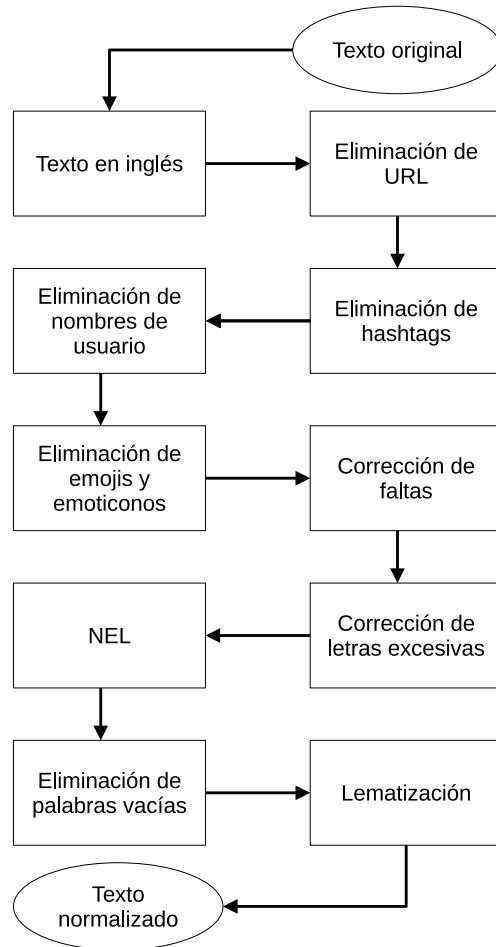


Figura 4.1: Esquema del proceso seguido en el preprocesamiento de los textos del *dataset*.

Capítulo 5

Resultados

La lista de profanidades que se ha empleado las categoriza en 6 grupos:

- **general:** profanidades sin ninguna connotación social o cultural concreta
- **LGBT:** relacionados con la identidad sexual
- **racial:** ataques o burlas contra la raza
- **religión:** sobre las creencias religiosas o la pertenencia a algún grupo religioso determinado
- **sexual:** todo lo referente a las preferencias sexuales de alguien
- **shock:** términos usados para causar conmoción o asco

En la figura 5.1 podemos ver un gráfico con el porcentaje de las profanidades de cada categoría presentes en el *dataset*.

Las profanidades generales están distribuidas de forma más desigual, con un 94 % de las profanidades de esta categoría presentes en Twitter y sólo 44 % en Mastodon. Para aquellas relacionadas con la comunidad LGBT, encontramos que el 32 % de ellas en Twitter y el 15 % en Mastodon. Y para aquellas que versan sobre la raza, 44 % en Twitter frente a un mero 8 % en Mastodon. En lo que respecta a la religión, encontramos el mismo porcentaje de palabras de esta categoría en ambas plataformas, un 71 %. El 51 % de las profanidades sexuales pueden encontrarse en Twitter, mientras que en Mastodon sólo están el 20 %. De las profanidades que causan shock, el 22 % estaban presentes en los textos de Twitter, pero ninguna en los de Mastodon.

También comprobamos cuántas publicaciones contenían al menos una profanidad, organizándolo por categoría y en total (véase el cuadro 5.1). También generamos una serie temporal para ver el porcentaje de los tweets y toots que contienen al menos una profanidad, la cual puede verse en la figura 5.2.

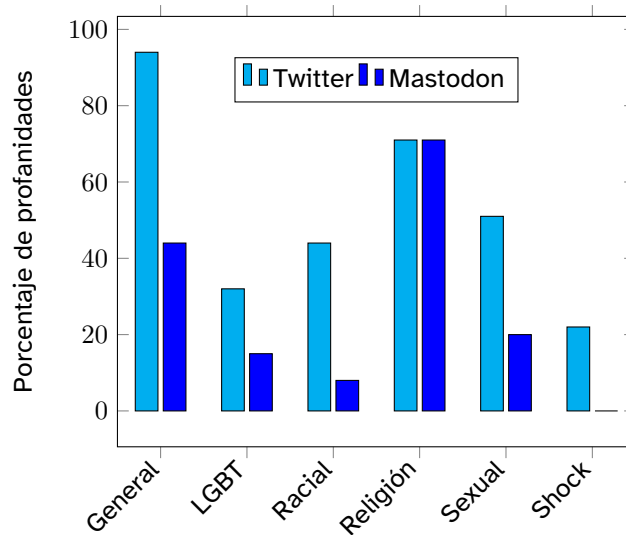


Figura 5.1: Porcentaje de profanidades de cada categoría presentes en Twitter y Mastodon

Cuadro 5.1: Publicaciones con al menos un profanidad de cada categoría y publicaciones con al menos una profanidad de cualquier categoría, en números absolutos y porcentajes del número total de publicaciones de cada plataforma.

Categoría		general	lgbt	racial	sexual	religión	shock	total
Twitter	N° pub.	87.235	14.267	1.429	86.282	27.138	7	209.166
	% pub.	2,28	0,37	0,04	2,26	0,71	0,00	5,47
Mastodon	N° pub.	330.288	21.614	2.736	531.527	80.097	0	929.260
	% pub.	2,23	0,15	0,02	3,59	0,54	0,00	6,28

El 18 % de palabras de la lista estaban presentes en las publicaciones de ambas plataformas. A su vez, el 31% de las profanidades estaban presentes de forma exclusiva en los textos de Twitter, mientras que el vocabulario de profanidades exclusivo de Mastodon se limita a una única palabra, “enby”.

En lo que respecta a la similitud coseno de las palabras presentes en ambos sitios (recordemos que pueden ir desde 0, menos similar, a 1, más similar), todos los términos exhibieron una similitud por debajo del 0,5. Las palabras con mayor similitud, en orden descendente, fueron “clover”, “rape”, “circlejerk” y “yellow-shower”, con puntuación de 0,492, 0,470, 0,432 y 0,41 respectivamente. En el fondo de la lista encontramos, de nuevo en orden descendente, “bullet-vibe”, “goddam”, “rectum” y “pornhub” con 0,039, 0,034, 0,026 y 0,023. La similitud media es de 0,177.

Encontramos un escenario similar con los términos cercanos de las palabras

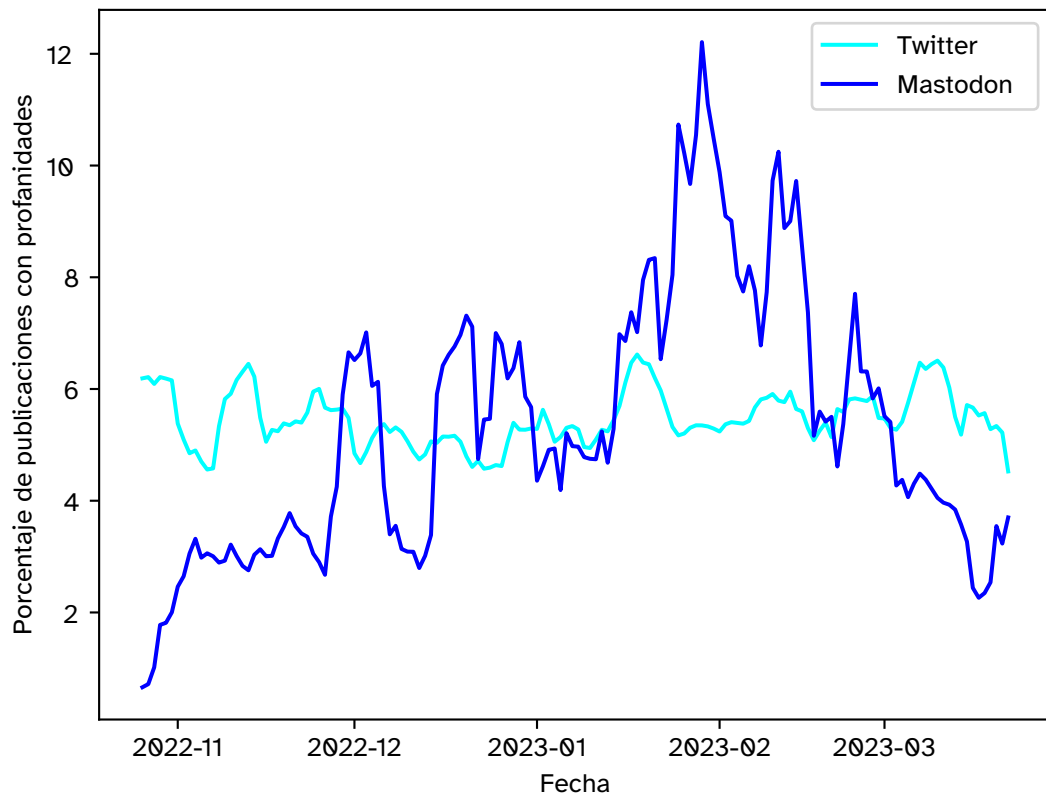


Figura 5.2: Serie temporal dónde se ve el porcentaje de publicaciones por día que contienen al menos una profanidad.

CAPÍTULO 5. RESULTADOS

polémicas. Sólo hubo un único término con una similitud superior a 0,5, “liberal”, con una puntuación de 0,513. Las siguientes cuatro palabras en similitud fueron “fraud” (0,431), “threat” (0,381), “stupid” (0,33) y “people” (0,33). La similitud coseno media en esta caso fue 0,161.

Habiendo presentado los resultados como tal, sería menester realizar un análisis de los mismos. En primer lugar, comprobaremos la distribución de las profanidades en las interacciones de cada red social (RQ1). Si prestamos atención a la figura 5.1 hay una conclusión clara: el vocabulario de profanidades en todas las categorías (a excepción de las religiosas) es mayor en Twitter que en Mastodon.

Empero, si comprobamos los porcentajes de tweet y toots que contienen profanidades en la tabla 5.1, vemos que son muy similares. Esto podría llevarnos a pensar que el volumen de mal vocabulario en ambas plataformas sigue patrones similares. Sin embargo, es necesario analizar las tendencias en la serie temporal de la figura 5.2 para ver una clara diferencia. Mientras que en Twitter tiende a haber un porcentaje de profanidades relativamente estable a lo largo del tiempo, en Mastodon el número va creciendo paulatinamente hasta alcanzar el pico en febrero de 2023, momento en el cual decrece hasta volver a niveles cercanos a los iniciales, aunque ligeramente más altos. Esto puede ser el resultado de la ola de nuevos usuarios de la “migración de Twitter”, pues ese nuevo influjo de usuarios se traería sus malos hábitos consigo, causando el repunte en vocabulario ofensivo. En un momento dado, esa ola de usuarios termina, y los recién llegados o bien abandonan Mastodon o se terminan acostumbrando a las normas sociales de la nueva plataforma, provocando el descenso de profanidades.

Esta idea de una diferencia en el comportamiento de los usuarios entre plataformas se ve reforzada por la riqueza del vocabulario de profanidades. Los usuarios de Twitter muestran un léxico de profanidades más amplio que los de Mastodon. Esta diferencia puede ser atribuida, en parte, al énfasis que el ecosistema de Mastodon pone en una moderación robusta. Los usuarios se distribuyen en servidores, cada uno con su propio código de conducta y moderadores. En las comunidades de Mastodon existe una preocupación genuina por mantener un buen comportamiento entre los usuarios, hasta el punto de que existan listas negras públicas de instancias “tóxicas” donde la moderación es prácticamente inexistente [GM21].

Esta idea gana fuerza cuando comprobamos las similitudes entre palabras. No sólo hay un importante número de palabras exclusivas a una sola de las plataformas, sino que el uso que se da a las que sí son comunes (RQ2) también varía. Esto sugiere que, aunque los periodistas siguen encontrándose con insultos y vocabulario degradante en sus interacciones en ambos sitios, el contexto, forma y frecuencia de dicho vocabulario varía entre plataformas.

Habiendo comprobado como aparecen las profanidades en cada plataforma (RQ1) y el uso que se da a las mismas (RQ2), podemos finalmente analizar cómo

los usuarios interactúan al discutir sobre temas controvertidos.

El método usado para examinar los temas polémicos pretende mostrar los cambios en el contexto cultural que los rodea. Esto permite identificar cambios no sólo en el uso de los propios términos, sino en todo el discurso que los rodea. Esto se vuelve especialmente importante cuando se discuten cuestiones que tienden a causar discusiones airadas. Analizando los resultados es evidente que las discusiones en Twitter y en Mastodon se desarrollan de forma diferente. La cuestión de si este cambio es positivo es objeto de futuras investigaciones. Empero, el actual estado de Twitter desde que lo comprara Elon Musk [Hi23] sugiere que alternativas como Mastodon puede que fomenten conversaciones más sanas.

Capítulo 6

Conclusiones y trabajo a futuro

En la introducción se subrayó cómo el modelo de actual de plataformas privadas y cerradas en los medios sociales presentaba riesgos tanto a los usuarios de dichas plataformas como a la sociedad en su conjunto, a la vez que subrayábamos las principales críticas a este modelo. A medida que estos problemas y preocupaciones se han ido acumulando, un número cada vez mayor de usuarios han recurrido a alternativas libres y más amigables con su comunidad.

Esta tendencia puede verse en la creciente popularidad de alternativas ya maduras técnicamente, como son las plataformas que conforman el Fediverso, especialmente desde que Elon Musk se hiciera con el control de Twitter en octubre de 2022.

Dado que los periodistas suelen ser de los primeros y principales objetivos del acoso en la red, y considerando que muchos de ellos se encuentran activos tanto en Twitter como en Mastodon, hemos tenido la oportunidad de contrastar cómo estas dos comunidades interactúan con los periodistas y comprobar si existen diferencias significativas entre las dos plataformas.

Tras analizar un nutrido *dataset* de tweets y toots, estamos en posición de arrojar algo de luz en la materia. En lo que respecta al uso de lenguaje ofensivo, los usuarios de Twitter tienen un vocabulario mucho más rico, con un abanico de profanidades más variado. Aunque el volumen total de profanidades es relativamente similar en ambos corpus, sospechamos que el influjo progresivo de usuarios de Twitter hacia Mastodon también ha introducido parte de ese vocabulario de Twitter. Además, cuando se compara directamente la forma en la que los usuarios utilizan esas mismas palabras, está claro que existe una diferencia: cuando los usuarios de Mastodon usan profanidades en sus conversaciones con periodistas, no lo hacen de la misma forma que los usuarios de Twitter.

CAPÍTULO 6. CONCLUSIONES Y TRABAJO A FUTURO

Observamos el mismo fenómeno cuando comprobamos el lenguaje que se emplea en las conversaciones entorno a cuestiones polémicas de índole social, económica o política. Es hasta cierto punto esperable que cuando un periodista trata determinadas cuestiones como son los refugiados o potenciales aumentos en el gasto social surjan debates y discusiones. Sin embargo, cuando comparamos los términos más cercanos a esas palabras contenciosas (los términos que conforman la conversación en torno a ellas), nos encontramos que también se usan de forma diferente en cada plataforma. Por ende, concluimos que existe una diferencia en la forma en que estas conversaciones se desarrollan entre los dos sitios.

Basándonos en los resultados de los análisis realizados sobre los datos recogidos, se vuelve evidente que existe una diferencia en cómo se trata en cada plataforma a los periodistas, objetivos habituales del acoso en Internet. No obstante, es necesario profundizar más en la materia, pues no nos es posible tomar estas diferencias como pruebas fehacientes de que Mastodon ofrece un entorno más “sano” en comparación con Twitter.

El trabajo futuro debería ahondar en determinar los cambios específicos en las interacciones de los usuarios de las dos plataformas, tratando de desentrañar si esos cambios tienen un impacto positivo o negativo. Además, se deben explorar métodos alternativos de recogida de datos en Mastodon, no sólo porque el sistema empleado en el presente trabajo es lento y no escalable, sino porque es necesario considerar las altas expectativas de privacidad que existen en el Fediverso. La investigación no se puede llevar a cabo a expensas de las personas de las que depende.

Capítulo 7

Dirección y gestión del proyecto

7.1 Planificación del proyecto

7.1.1 Identificación de interesados

La lista de interesados correspondería sería la siguiente:

1. Ángel García Menéndez (proyectante)
2. Daniel Gayo Avello (tutor)
3. Periodistas investigados
4. Usuarios de los cuales se descarguen los datos
5. Administradores de las instancias de Mastodon
6. Editor de la revista receptora del artículo
7. Revisores del artículo
8. Miembros del tribunal evaluador

7.1.2 OBS y PBS

El OBS puede encontrarse en la figura 7.1. Al tratarse el proyecto de un trabajo de final de máster sólo cuenta con dos elementos, el tutor, Daniel Gayo Avello y el proyectante, Ángel García Menéndez.

El PBS o estructura de desglose de producto se plantea a continuación. Para cada tarea planteada en la EDT se genera un producto.

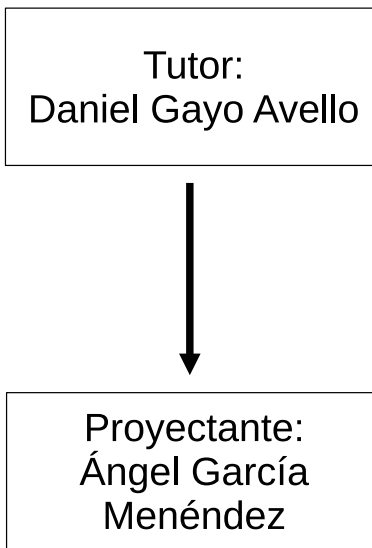


Figura 7.1: OBS u organigrama del proyecto

1. Planificación del proyecto
 - 1.1. Lista de interesados
 - 1.2. Lista de requisitos
 - 1.3. Alcance del proyecto
 - 1.4. OBS
 - 1.5. PBS
 - 1.6. EDT
 - 1.7. Planificación temporal
 - 1.7.1. Actividades del proyecto
 - 1.7.2. Ordenación de las actividades
 - 1.7.3. Lista de recursos del proyecto
 - 1.7.4. Duración de las actividades
 - 1.7.5. Cronograma
 - 1.8. Planificación económica
 - 1.8.1. Costes estimados del proyecto
 - 1.8.2. Presupuesto inicial del proyecto
 - 1.9. Planificación de la gestión de riesgos
 - 1.9.1. Lista de riesgos
 - 1.9.2. Análisis cualitativo
 - 1.9.3. Análisis cuantitativo
 - 1.9.4. Lista de respuestas
 - 1.10. Acta de proyecto
 - 1.11. Acta de reunión de aprobación
2. Estado del arte
 - 2.1. Acta de reunión de preparación de estado del arte
 - 2.2. Revisión de trabajos de Descripción de red social
 - 2.3. Revisión de trabajos de Espiral del silencio
 - 2.4. Revisión de trabajos de Diferencias de discurso
 - 2.5. Revisión de trabajos de Dinámicas de viralidad
 - 2.6. Revisión de trabajos de Hostigamiento a periodistas
 - 2.7. Revisión de trabajos de Ética

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

- 2.8. Acta de reunión de valoración del estado del arte
- 3. Datos del proyecto
 - 3.1. Acta de reunión de preparación de obtención de datos
 - 3.2. Lista de periodistas
 - 3.3. Código de descarga de Twitter
 - 3.4. Datos en bruto de Twitter
 - 3.5. Código de descarga de Mastodon
 - 3.6. Datos en bruto de Mastodon
 - 3.7. Acta de reunión de valoración de obtención de datos
- 4. Datos procesados
 - 4.1. Acta de reunión de preparación de procesamiento de datos
 - 4.2. Datos normalizados
 - 4.3. Datos anonimizados
 - 4.4. Análisis sobre los datos
 - 4.5. Informe de resultados
 - 4.6. Acta de reunión de valoración de procesamiento de datos
- 5. Artículo científico
 - 5.1. Acta de reunión de preparación de elaboración del artículo
 - 5.2. Apartado de Metodología
 - 5.3. Apartado de Resultados
 - 5.4. Apartado de Conclusiones
 - 5.5. Apartado de Estado del arte
 - 5.6. Apartado de Introducción
 - 5.7. Apartado de Resumen y Título
 - 5.8. Acta de reunión de valoración de elaboración del artículo
 - 5.9. Artículo corregido
 - 5.10. Documentación acreditativa del envío
- 6. Documentación de cierre de proyecto
 - 6.1. Memoria del proyecto
 - 6.2. Análisis de la gestión y dirección
 - 6.2.1. Planificación final

6.2.2. Presupuesto final

6.2.3. Informe de riesgos

6.2.4. Informe de lecciones aprendidas

6.3. Documentación acreditativa del envío en SIES

7.1.3 Planificación inicial (EDT)

En esta sección se desgrena la planificación inicial planteada para el proyecto. Puede encontrarse también desglosada en el fichero `planificacion_inicial.mpp` adjunto. Se establecen 6 grandes grupos de tareas, habiendo para cada uno una tabla con las tareas que lo componen y un diagrama de Gantt para visualizar la planificación:

- **Tarea 1:** Planificación del proyecto (véanse las tareas en el cuadro 7.1 y el diagrama de Gantt en la figura 7.2)
- **Tarea 24:** Estudio del estado del arte (véanse las tareas en el cuadro 7.5 y el diagrama de Gantt en la figura 7.6)
- **Tarea 33:** Obtención de datos (véanse las tareas en el cuadro 7.6 y el diagrama de Gantt en la figura 7.7)
- **Tarea 41:** Procesamiento de los datos (véanse las tareas en el cuadro 7.7 y el diagrama de Gantt en la figura 7.8)
- **Tarea 48:** Elaboración del artículo (véanse las tareas en el cuadro 7.8 y el diagrama de Gantt en la figura 7.9)
- **Tarea 58:** Cierre del proyecto (véanse las tareas en el cuadro 7.9 y el diagrama de Gantt en la figura 7.10)

Esta primera planificación establecía como fecha de inicio del proyecto el 1 de marzo de 2023, y estima que se finalizará el 7 de julio de 2023, con una duración de 226 días. Se establece un calendario de trabajo a media jornada (4 horas), por encontrarse el proyectante compaginando el proyecto con su trabajo de profesor en la Universidad de Oviedo.

En el caso de esta primera fase de Planificación del proyecto, se incluyen también desglosadas las tareas de Planificación temporal (tarea con identificador 08, véanse el cuadro 7.2 y la figura 7.3), Planificación económica (tarea con identificador 14, véanse el cuadro 7.3 y la figura 7.4) y Planificación de la gestión de riesgos (tarea con identificador 17, véanse el cuadro 7.4 y la figura 7.5).

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.1: Desglose de las tareas de correspondientes a la Planificación del proyecto (tarea con identificador 01)

Id	Nombre	Antecesora	Duración	Comienzo
02	Identificación de interesados	N/A	4 horas	2023-03-01
03	Captación de requisitos	2	5 días	2023-03-01
04	Definición de alcance	3	2 días	2023-03-06
05	Elaboración del OBS	4	2 horas	2023-03-27
06	Elaboración del PBS	5	2 horas	2023-03-27
07	Elaboración de la EDT	6	3 horas	2023-03-08
08	Planificación temporal	7	5,75 días	2023-03-08
14	Planificación económica	8	4 días	2023-03-13
17	Planificación de gestión de riesgos	14	5,75 días	2023-03-16
22	Elaboración del acta de proyecto	17	2 días	2023-03-21
23	Reunión de aprobación del proyecto	22	4 horas	2023-03-22

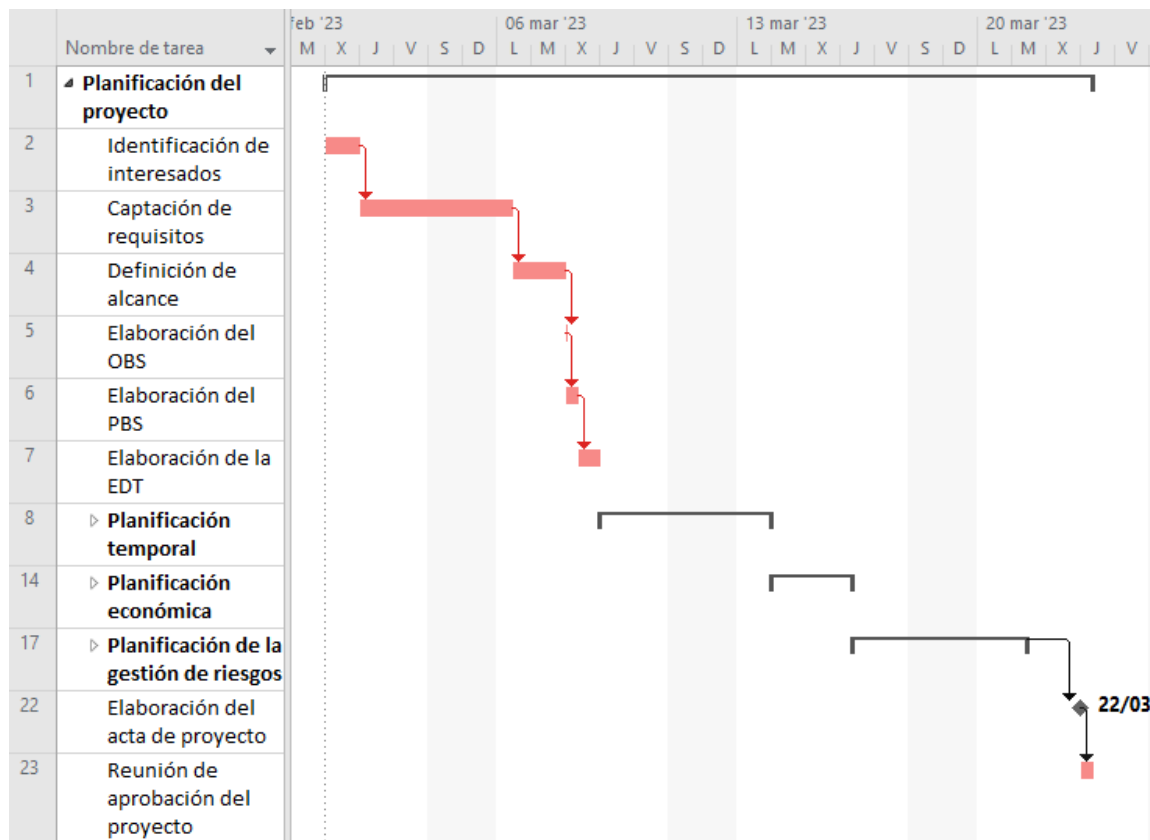


Figura 7.2: Diagrama de Gantt de las tareas correspondientes a la Planificación del proyecto (tarea con identificador 01)

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.2: Desglose de las tareas de correspondientes a la Planificación temporal (tarea con identificador 08)

Id	Nombre	Antecesora	Duración	Comienzo
9	Definición de actividades	7	3 horas	2023-03-08
10	Secuenciación de actividades	9	2 horas	2023-03-09
11	Estimación de recursos de las actividades	10	2 horas	2023-03-09
12	Estimación de duración de las actividades	11	2 días	2023-03-09
13	Desarrollo del cronograma	12	2 días	2023-03-10

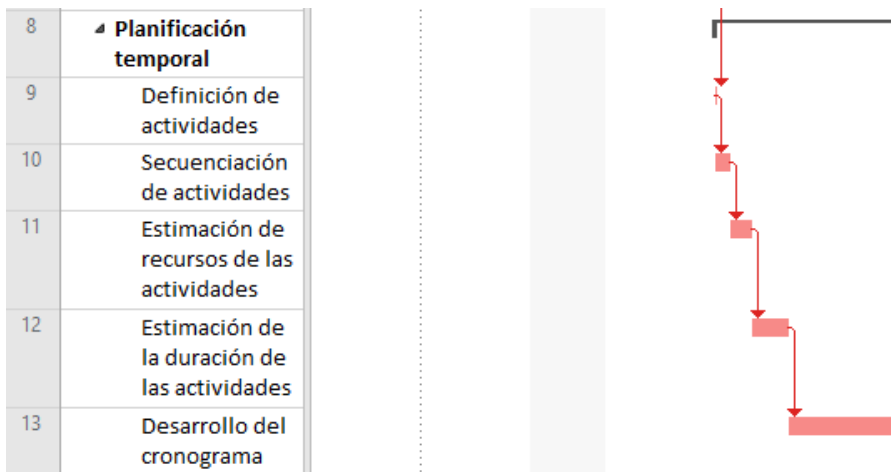


Figura 7.3: Diagrama de Gantt de las tareas correspondientes a la Planificación temporal (tarea con identificador 08)

Cuadro 7.3: Desglose de las tareas de correspondientes a la Planificación económica (tarea con identificador 14)

Id	Nombre	Antecesora	Duración	Comienzo
15	Estimación de costes	8	2 días	2023-03-13
16	Elaboración de presupuestos	15	2 días	2023-03-15

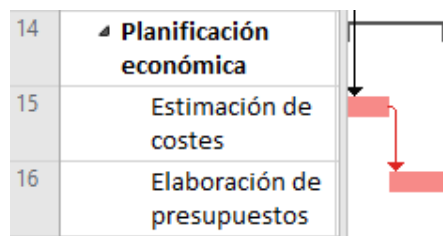


Figura 7.4: Diagrama de Gantt de las tareas correspondientes a la Planificación económica (tarea con identificador 14)

Cuadro 7.4: Desglose de las tareas de correspondientes a la Planificación de gestión de riesgos (tarea con identificador 17)

Id	Nombre	Antecesora	Duración	Comienzo
18	Identificación de riesgos	14	2 días	2023-03-16
19	Análisis cualitativo	18	3 horas	2023-03-17
20	Análisis cuantitativo	19	4 horas	2023-03-17
21	Planificación de respuestas	20	2 días	2023-03-20

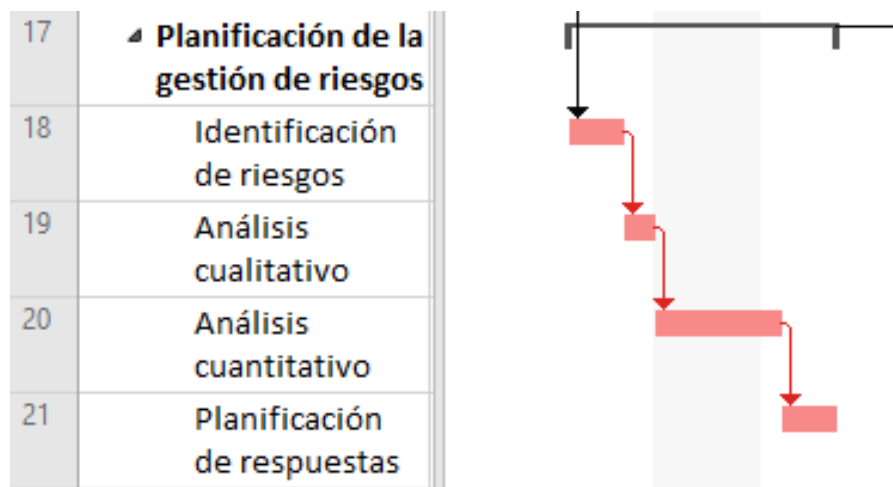


Figura 7.5: Diagrama de Gantt de las tareas correspondientes a la Planificación de gestión de riesgos (tarea con identificador 17)

Cuadro 7.5: Desglose de las tareas de correspondientes al Estudio del estado del arte (tarea con identificador 24)

Id	Nombre	Antecesora	Duración	Comienzo
25	Reunión de preparación del estado del arte	1	4 horas	2023-03-23
26	Descripción de la red social	25	7 días	2023-03-23
27	Espiral del silencio	26	5 días	2023-03-29
28	Diferencias de discurso	27	5 días	2023-04-03
29	Dinámicas de viralidad	28	7 días	2023-04-06
30	Hostigamiento a periodistas	29	3,5 días	2023-04-12
31	Ética en investigaciones en redes sociales	30	5 días	2023-04-19
32	Reunión de valoración de estado el arte	31	4 horas	2023-04-19

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

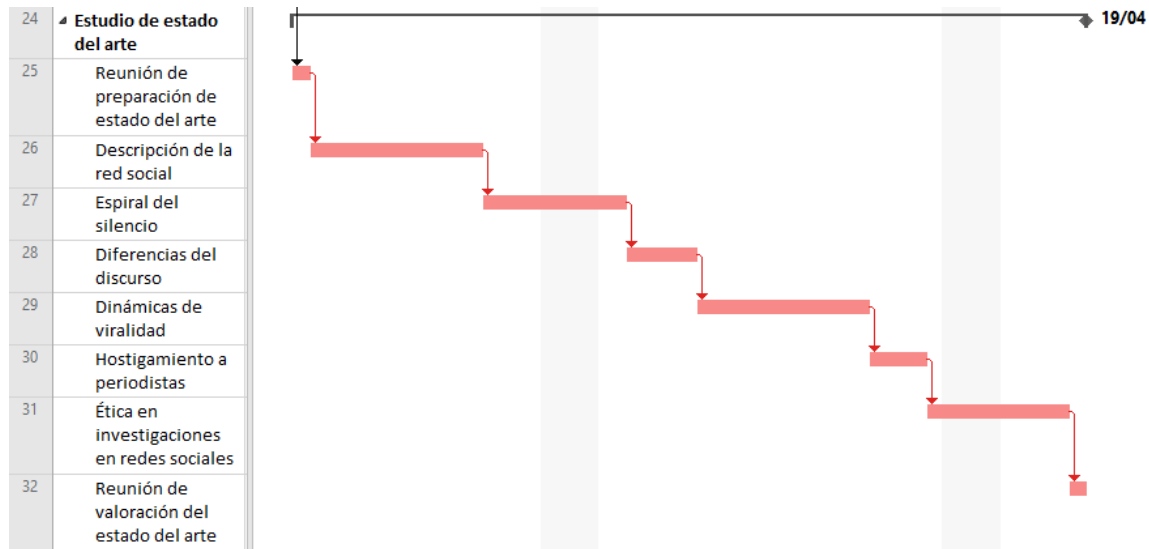


Figura 7.6: Diagrama de Gantt de las tareas correspondientes al Estudio del estado del arte (tarea con identificador 24)

Cuadro 7.6: Desglose de las tareas de correspondientes a la Obtención de datos (tarea con identificador 33)

Id	Nombre	Antecesora	Duración	Comienzo
34	Reunión de preparación de obtención de datos	24	2 días	2023-04-19
35	Identificación de periodistas	34	6 días	2023-04-21
36	Desarrollo del código de descarga de Twitter	35	4 días	2023-04-26
37	Descarga de datos de Twitter	36	7 días	2023-04-28
38	Desarrollo del código de descarga de Mastodon	37	7 días	2023-05-04
39	Descarga de datos de Mastodon	38	10 días	2023-05-10
40	Reunión de valoración de obtención de datos	39	1,75 días	2023-05-23

Cuadro 7.7: Desglose de las tareas de correspondientes al Procesamiento de datos (tarea con identificador 41)

Id	Nombre	Antecesora	Duración	Comienzo
42	Reunión de preparación de procesamiento de datos	33	4 horas	2023-05-19
43	Limpieza de los datos	42	7 días	2023-05-22
44	Anonimización de los datos	43	1 día	2023-05-26
45	Análisis de los datos	44	5 días	2023-05-26
46	Elaboración de informe de resultados	45	2 días	2023-05-31
47	Reunión de valoración de procesamiento de datos	46	4 horas	2023-06-01

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

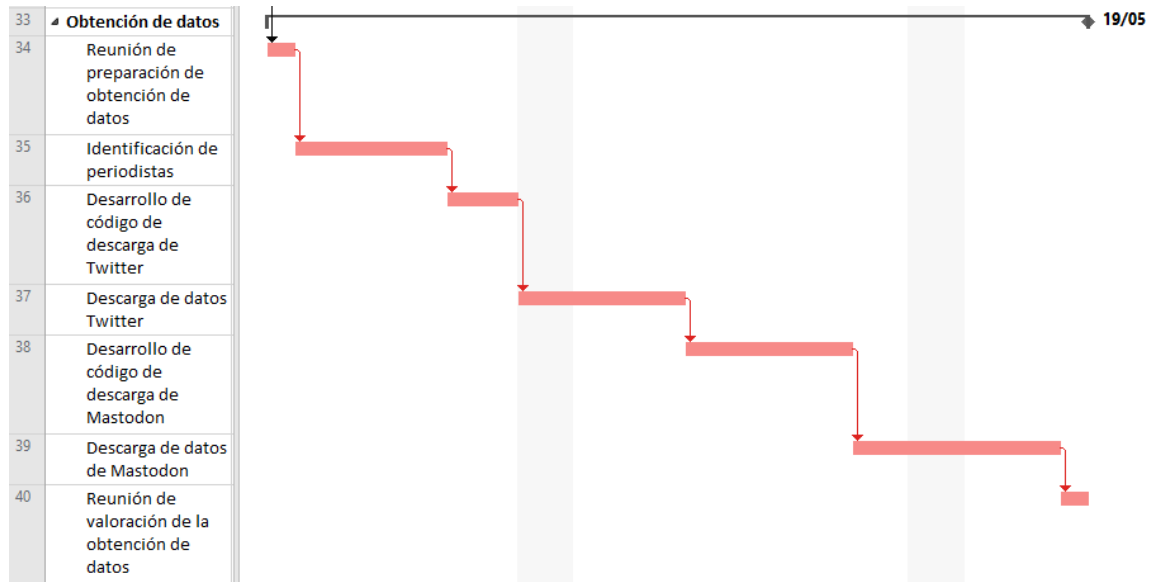


Figura 7.7: Diagrama de Gantt de las tareas correspondientes a la Obtención de datos (tarea con identificador 33)

Cuadro 7.8: Desglose de las tareas de correspondientes a la Elaboración del artículo (tarea con identificador 48)

Id	Nombre	Antecesora	Duración	Comienzo
49	Reunión de preparación del artículo	41	4 horas	2023-06-02
50	Redacción de la metodología	49	1,75 días	2023-06-02
51	Redacción de los resultados	50	3 días	2023-06-05
52	Redacción de las conclusiones	51	3,5 días	2023-06-07
53	Redacción del estado del arte	52	4 días	2023-06-09
54	Redacción de la introducción	53	3,5 días	2023-06-13
55	Redacción del resumen y título	54	3,5 días	2023-06-15
56	Reunión de valoración del artículo	55	4 horas	2023-06-19
57	Envío del artículo	56	2 horas	2023-06-20

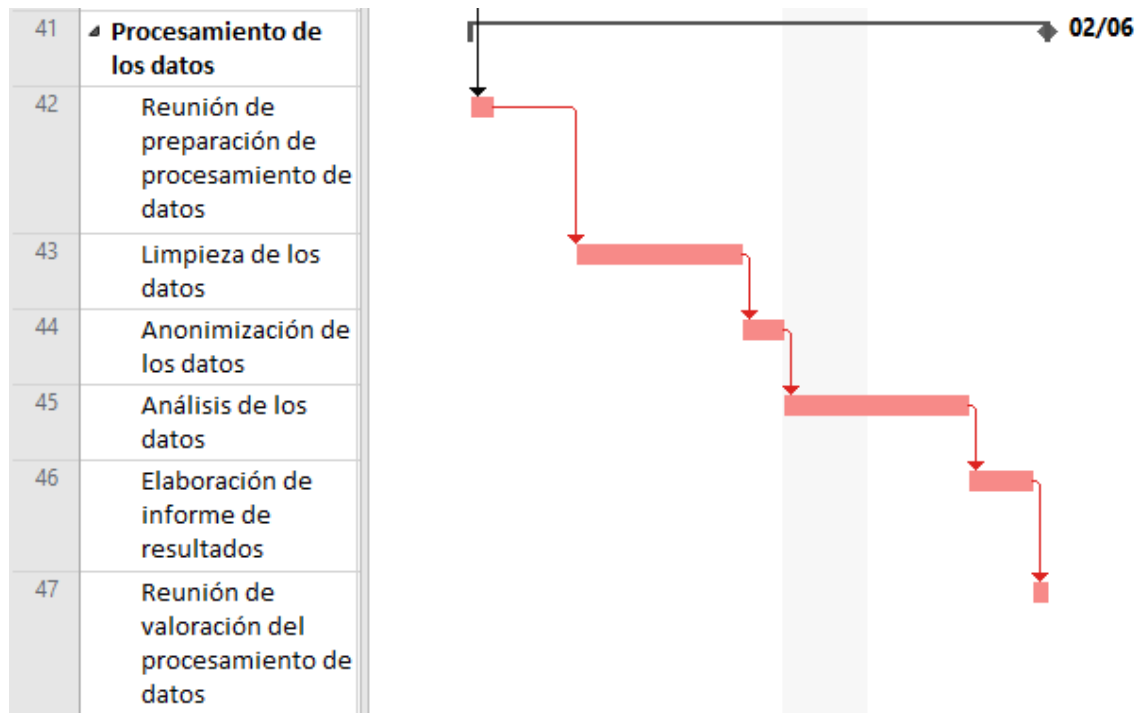


Figura 7.8: Diagrama de Gantt de las tareas correspondientes al Procesamiento de datos (tarea con identificador 41)

Cuadro 7.9: Desglose de las tareas de correspondientes al Cierre de Proyecto (tarea con identificador 58). Las tareas 61, 62, 63 y 64 son hijas de la tarea 60.

Id	Nombre	Antecesora	Duración	Comienzo
59	Traducción de la memoria	48	5 días	2023-06-20
60	Análisis de dirección y gestión	N/A	17 días	2023-06-23
61	Elaboración de la planificación final	59	5 días	2023-06-23
62	Elaboración de presupuesto final	61	4 días	2023-06-28
63	Elaboración de informe final de riesgos	62	7 días	2023-06-30
64	Elaboración de informe de lecciones aprendidas	63	1 día	2023-07-06
65	Envío de la memoria	60	1 hora	2023-07-07

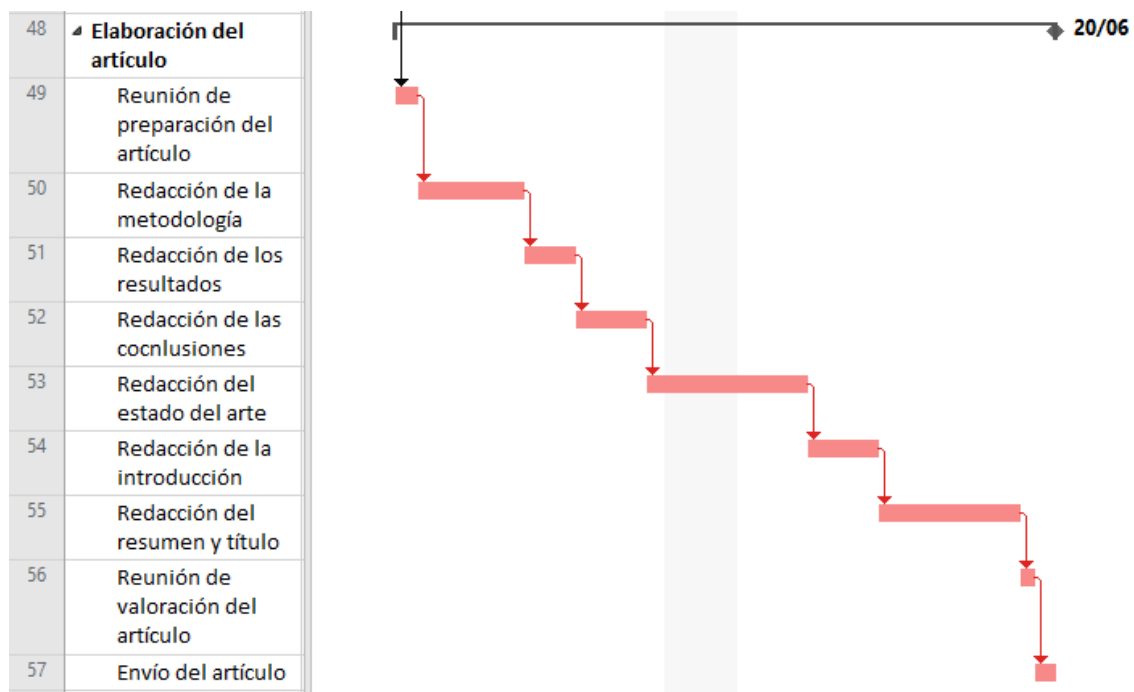


Figura 7.9: Diagrama de Gantt de las tareas correspondientes a la Elaboración del artículo (tarea con identificador 48)

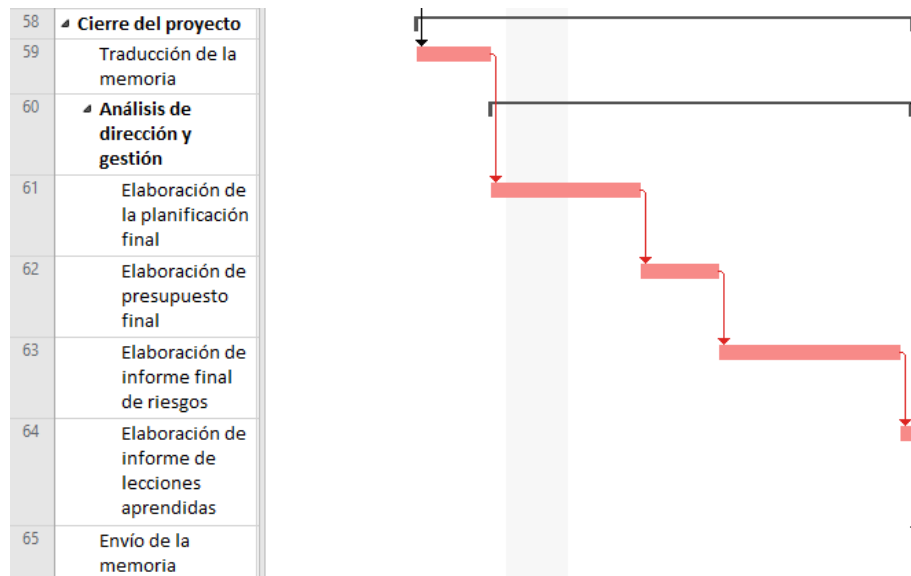


Figura 7.10: Diagrama de Gantt de las tareas correspondientes al Cierre de Proyecto (tarea con identificador 58)

7.1.4 Gestión de riesgos

7.1.4.1 Plan de gestión de riesgos

El plan de gestión de riesgos se encuentra desglosado en el anexo A, dónde se expone la metodología empleada, la clasificación de los riesgos, así como las técnicas usadas para analizarlos.

7.1.4.2 Identificación de los riesgos

Se han identificado los siguientes riesgos en el proyecto:

1. No se consiguen suficientes usuarios como para hacer experimentos significativos
2. Limitaciones de la API de Mastodon
3. Retraso de las tareas relacionadas con el Estado del Arte
4. Pérdida de Twitter como fuente de datos
5. Conflicto en los términos de uso de las instancias de Mastodon

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

6. Conflicto con las expectativas de privacidad de los usuarios
7. Los experimentos considerados a priori no arrojan resultados significativos
8. Aparece una investigación similar
9. Artículo es rechazado por ser el primero del autor
10. Indisposición del proyectante
11. Pérdida del *dataset*
12. Retrasos en la redacción del artículo por falta de experiencia
13. Cambios en la regulación de la protección de datos
14. Retrasos por la falta de hardware para procesar el *dataset*
15. Recursos desarrollados por parte de otros investigadores para investigaciones similares no están disponibles

7.1.4.3 Registro de riesgos

En el cuadro 7.10 podemos ver el registro de riesgos. Cada riesgo se identifica con el número asignado en 7.1.4.2, y se analiza de acuerdo a los criterios expuestos en el Plan de Gestión de Riesgos (ver apéndice A). En el cuadro 7.1.4.3 se pueden comprobar las respuestas a cada riesgo.

Cuadro 7.10: Registro de los riesgos, dónde se expone la probabilidad y el impacto que se les anticipa en el proyecto.

ID	Prob.	Impacto				Impacto total
		Presupuesto	Planificación	Alcance	Calidad	
1	Media	Muy bajo	Alto	Alto	Medio	0,28
2	Media	Muy bajo	Alto	Bajo	Medio	0,28
3	Media	Muy bajo	Medio	Medio	Medio	0,15
4	Media	Muy bajo	Alto	Alto	Medio	0,28
5	Alta	Muy bajo	bajo	Medio	Bajo	0,21
6	Baja	Bajo	Bajo	Muy alto	Bajo	0,27
7	Media	Alto	Muy alto	Alto	Alto	0,45
8	Baja	Muy bajo	Bajo	Medio	Alto	0,17
9	Baja	Muy bajo	Muy bajo	Alto	Alto	0,17
10	Media	Medio	Alto	Medio	Medio	0,28
11	Baja	Bajo	Alto	Bajo	Medio	0,17
12	Alta	Medio	Alto	Bajo	Medio	0,39
13	Baja	Medio	Medio	Alto	Alto	0,17
14	Media	Alto	Alto	Medio	Bajo	0,28
15	Alta	Bajo	Medio	Bajo	Medio	0,21

Cuadro 7.11: Respuestas para cada riesgo, con el tipo de respuesta y la descripción de la misma

ID	Tipo de respuesta	Respuesta
1	Mitigación	Se reestructurarán los experimentos para adaptarlos a la cantidad de datos recogidos
2	Mitigación	Se buscarán métodos de obtención de datos alternativos a la API, como el raspado web
3	Asunción	Se escapa a nuestro control el si la literatura será más o menos accesible
4	Asunción	Dado el actual devenir de la empresa, escapa del todo a nuestro control
5	Eliminación	Una vez descargados los textos, se comprobarán manualmente las instancias afectadas para filtrar aquellas con prohibiciones expresas
6	Mitigación	No se tomarán datos de aquellos usuarios que lo indiquen de forma estándar en su perfil
7	Mitigación	Se reestructurarán los experimentos en caso de que se de el riesgo
8	Asunción	Escapa del todo a nuestro control
9	Mitigación	En caso de que el riesgo se manifieste, será el tutor el que envíe el artículo a la revista
10	Asunción	Escapa del todo a nuestro control
11	Eliminación	Una vez completada la descarga y el preprocesamiento del dataset, este se guardará en la nube
12	Asunción	Al ser el primer artículo que escribe el proyectante, se ha de asumir que llevará más tiempo que a un profesional consagrado
13	Asunción	Escapa a nuestro control
14	Mitigación	Se buscarán formas alternativas de procesar el dataset que no sean tan costosas en cuanto a hardware
15	Asunción	No se puede controlar qué hagan otros investigadores

7.1.5 Presupuesto inicial

En esta sección se detalla el presupuesto planteado en el inicio del proyecto. Puede encontrarse completamente desglosado en el fichero `presupuesto_inicial.xlsx` adjunto.

Se trata de un presupuesto exclusivamente de costes. La razón para ello es que el proyecto se plantea como una investigación interna, ya que por su carácter social y no eminentemente práctico sería difícil “venderlo” de forma alguna. Por ende, en lo sucesivo se entenderá el proyecto como una investigación de carácter interno llevada a cabo en el contexto de empresa presentado en la asignatura de Dirección y Planificación de Proyectos Web. El contexto económico de la empresa, especialmente de los costes de mano de obra, puede encontrarse en el fichero `contexto_empresa.xlsx`.

Lo que más nos interesa es el coste del investigador, situado en 28.000 € anuales. Teniendo en cuenta que las horas de trabajo anuales en España se encuentran en torno a las 1820¹, establecemos que el coste por hora es de 15,38 €. Por lo demás se entiende que el equipo en el que se investiga, así como los que se usaron para la descarga del *dataset*, y las licencias del MS Project y Windows 10 (el resto del software empleado es libre) son amortizaciones de la empresa, y por ende no se reflejan en el presupuesto.

El presupuesto se encuentra dividido en 7 partidas, una para cada fase de las descritas en la planificación de la sección 7.1.3, y la séptima con gastos misceláneos, principalmente las reservas de contingencia para los riesgos monitorizados en la sección 7.2.3:

- **Partida 1 - Planificación del proyecto** (véase cuadro 7.13)
- **Partida 2 - Estudio del estado del arte** (véase cuadro 7.14)
- **Partida 3 - Obtención de datos** (véase cuadro 7.15)
- **Partida 4 - Procesamiento de los datos** (véase cuadro 7.16)
- **Partida 5 - Elaboración del artículo** (véase cuadro 7.17)
- **Partida 6 - Cierre de proyecto** (véase cuadro 7.18)
- **Partida 7 - Otros** (véase cuadro 7.19)

El presupuesto de costes resumido puede verse en el cuadro 7.12, teniendo el proyecto un coste inicial de 9.997,00 €. El presupuesto final puede encontrarse en la sección 7.3.3.

¹Fuente: <https://www.expansion.com/diccionario-juridico/jornada-laboral.html>

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.12: Presupuesto de costes resumido

Cod.	Partida	Total (€)
01	Planificación del proyecto	1.737,94
02	Estudio del estado del arte	2.122,44
03	Obtención de datos	1.276,54
04	Procesamiento de los datos	1.045,84
05	Elaboración del artículo	1.338,06
06	Cierre de proyecto	1.368,82
07	Otros	1.107,36
TOTAL		9.997,00

Cuadro 7.13: Desglose de la partida 1

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total (€)
01		Identificación de interesados	4	horas	15,38		61,52
02		Captación de requisitos	20	horas	15,38		307,60
03		Definición de alcance	8	horas	15,38		123,04
04		Elaboración del OBS	2	horas	15,38		30,76
05		Elaboración del PBS	2	horas	15,38		30,76
06		Elaboración de la EDT	3	horas	15,38		46,14
07		Planificación temporal					353,74
	001	Definición de actividades	3	horas	15,38	46,14	
	002	Secuenciación de actividades	2	horas	15,38	30,76	
	003	Estimación de recursos de las actividades	2	horas	15,38	30,76	
	004	Estimación de la duración de las actividades	8	horas	15,38	123,04	
	005	Desarrollo del cronograma	8	horas	15,38	123,04	
08		Planificación económica					246,08
	001	Estimación de costes	8	horas	15,38	123,04	
	002	Elaboración de presupuestos	8	horas	15,38	123,04	
09		Planificación de la gestión de riesgos					353,74
	001	Identificación de riesgos	8	horas	15,38	123,04	
	002	Análisis cualitativo	3	horas	15,38	46,14	
	003	Análisis cuantitativo	4	horas	15,38	61,52	
	004	Planificación de respuestas	8	horas	15,38	123,04	
10		Elaboración del acta de proyecto	8	horas	15,38		123,04
11		Reunión de aprobación del proyecto	4	horas	15,38		61,52
						TOTAL	1.737,94

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.14: Desglose de la partida 2

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación de estado del arte	4	horas	15,38	61,52
02	Descripción de la red social	28	horas	15,38	430,64
03	Espiral del silencio	20	horas	15,38	307,60
04	Diferencias del discurso	20	horas	15,38	307,60
05	Dinámicas de viralidad	28	horas	15,38	430,64
06	Hostigamiento a periodistas	14	horas	15,38	215,32
07	Ética en investigaciones en redes sociales	20	horas	15,38	307,60
08	Reunión de valoración del estado del arte	4	horas	15,38	61,52
				TOTAL	2.122,44

Cuadro 7.15: Desglose de la partida 3

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación de obtención de datos	8	horas	15,38	123,04
02	Identificación de periodistas	24	horas	15,38	369,12
03	Desarrollo de código de descarga de Twitter	16	horas	15,38	246,08
04	Descarga de datos Twitter	28	horas	0,00	0,00
05	Desarrollo de código de descarga de Mastodon	28	horas	15,38	430,64
06	Descarga de datos de Mastodon	40	horas	0,00	0,00
07	Reunión de valoración de la obtención de datos	7	horas	15,38	107,66
				TOTAL	1.276,54

Cuadro 7.16: Desglose de la partida 4

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación de procesamiento de datos	4	horas	15,38	61,52
02	Limpieza de los datos	28	horas	15,38	430,64
03	Anonimización de los datos	4	horas	15,38	61,52
04	Análisis de los datos	20	horas	15,38	307,60
05	Elaboración de informe de resultados	8	horas	15,38	123,04
06	Reunión de valoración del procesamiento de datos	4	horas	15,38	61,52
				TOTAL	1.045,84

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.17: Desglose de la partida 5

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación del artículo	4	horas	15,38	61,52
02	Redacción de la metodología	7	horas	15,38	107,66
03	Redacción de los resultados	12	horas	15,38	184,56
04	Redacción de las conclusiones	14	horas	15,38	215,32
05	Redacción del estado del arte	16	horas	15,38	246,08
06	Redacción de la introducción	14	horas	15,38	215,32
07	Redacción del resumen y título	14	horas	15,38	215,32
08	Reunión de valoración del artículo	4	horas	15,38	61,52
09	Envío del artículo	2	horas	15,38	30,76
TOTAL					1.338,06

Cuadro 7.18: Desglose de la partida 6

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total
01		Traducción y adaptación de la memoria	20	horas	15,38		307,60
02		Análisis de dirección y gestión					1.045,84
	001	Elaboración de la planificación final	20	horas	15,38	307,60	
	002	Elaboración de presupuesto final	16	horas	15,38	246,08	
	003	Elaboración del informe de riesgos	28	horas	15,38	430,64	
	004	Elaboración del informe de lecciones aprendidas	4	horas	15,38	61,52	
03		Envío de la memoria	1	horas	15,38		15,38
TOTAL							1.368,82

Cuadro 7.19: Desglose de la partida 7

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total (€)
01		Reservas de contingencia					624,24
	001	Reestructuración de experimentos 1	36	horas	8,67	312,12	
	002	Reestructuración de experimentos 2	36	horas	8,67	312,12	
	003	Nueva metodología de descarga	24	horas	8,67	208,08	
TOTAL							1.107,36

7.2 Ejecución del proyecto

7.2.1 Plan de seguimiento de la planificación

Toda la monitorización del proyecto se llevará a cabo en dos puntos principales. El primero son las reuniones periódicas mantenidas con el tutor para monitorizar el proyecto, habiendo una cada semana los lunes. El segundo son las reuniones que se establecen al principio y final de cada fase de las planteadas en la sección 7.1.3, donde se pormenoriza qué se va a hacer en cada fase y al final se analiza lo hecho.

Se establecen dentro del proyecto 5 líneas base:

- **1 de marzo de 2023:** Línea base inicial, marcada por el comienzo del proyecto.
- **19 de mayo de 2023:** Fecha de finalización de la fase de obtención de datos. En este punto del proyecto se debería haber obtenido el *dataset* con los textos recabados de ambas plataformas.
- **2 de junio de 2023:** Fecha de finalización de la fase de procesamiento de datos. Los experimentos deberían estar concluidos y debería poder darse respuesta al riesgo 07 (véase sección 7.2.3.5).
- **20 de junio de 2023:** Fecha de finalización de la elaboración del artículo. En este punto el artículo debería haberse redactado y enviado ya a la revista.
- **7 de julio de 2023:** Línea base final, marcada por la finalización del proyecto y el envío de la memoria.

7.2.2 Bitácora de incidencias

En el cuadro 7.20 se encuentra la bitácora que recoge las principales incidencias encontradas durante el desarrollo del proyecto.

Lo más destacable son las incidencias 02 y 09. La primera es manifestación del riesgo 02 (ver sección 7.2.3.2), pues al ser la primera vez que se empleaba la API no se tenía ese conocimiento a priori. La segunda corresponde al riesgo 07 (ver sección 7.2.3.5), los experimentos planteados originalmente, basados en análisis de sentimiento, no arrojaron resultados de valor. Por ello se tomó la decisión expresada en la incidencia 10, de posponer la finalización del proyecto y reenfocar los experimentos a lo que finalmente se hizo (ver sección 4.3).

Cuadro 7.20: Bitácora de incidencias del proyecto

Id	Fecha	Descripción
01	2023-05-08	Se detecta un problema con la estructura federada de Mastodon, pues cada instancia tiene diferentes identificadores internos
02	2023-05-10	Debido a que sólo se puede solicitar un único post con la API, se tarde mucho en la descarga
03	2023-06-12	Se realiza una reunión para el análisis. Se decide trabajar con la salida de un NEL semimanual
04	2023-06-15	No se tiene acceso a recursos del artículo que se toma de referencia para la normalización
05	2023-06-17	Algunos de los ficheros que se han descargado son tan grandes que no se pueden cargar por entero en la memoria
06	2023-06-20	Se detecta gran cantidad de ruido en la lista de entidades obtenida
07	2023-06-29	Se detecta que podría haber sido necesario descargar citas a tweets, no sólo respuestas
08	2023-07-03	Se detecta que las URL no se eliminaron correctamente en los posts de Mastodon
09	2023-07-08	Los análisis propuestos no arrojan resultados significativos
10	2023-07-10	Se tiene una reunión para posponer la entrega del proyecto hasta noviembre
11	2023-10-06	El primer borrador del artículo resulta muy mejorable y es necesario realizar numerosas correcciones
12	2023-13-10	Se descubre que la retroalimentación de las entregas de DGPW del curso anterior son inaccesibles

7.2.3 Seguimiento de riesgos

7.2.3.1 Riesgo 01

Nombre: No se consiguen suficientes publicaciones para hacer experimentos significativos

Descripción: Para poder realizar análisis de cierta enjundia, es necesario con un mínimo número de publicaciones. Puede ser que la hipótesis de los periodistas como “canario de mina” no sea acertada.

Categorías: requisitos, calidad, usuarios

Probabilidad	Impacto				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy bajo	Alto	Alto	Medio	0,28	Mitigación
Probabilidad final	Impacto final				Impacto total	Respuesta efectiva
	Presupuesto	Planificación	Alcance	Calidad		
Baja	Muy bajo	Alto	Alto	Medio	0,28	No se produjo

Riesgos derivados: N/A

Riesgo residual: N/A

Plan de contingencia: En caso de que no se llegue a un mínimo de publicaciones en el dataset de cada red social, que se establece en 500.000 por cada una, será necesario reestructurar los experimentos de la fase de experimentación para adaptarlos a la cantidad de información con la que se cuenta.

Planificación temporal de las contingencias: El plan de contingencia planteado se ejecutaría entre el final la fase de recogida de datos y el comienzo de la fase de análisis.

Presupuesto para contingencias

Ítem	Concepto	Asignación (€)
1	Reestructuración de experimentos	553,68
	TOTAL	553,68

Monitorización: No se ha producido el riesgo.

7.2.3.2 Riesgo 02

Nombre: Limitaciones de la API de Mastodon

Descripción: Mastodon es una red social federada, similar al correo electrónico, con usuarios registrados en diferentes servidores que interactúan entre sí. Esta particularidad, sumada al hecho de no haber trabajado nunca con su API oficial, puede suponer un riesgo del proyecto.

Categorías: tecnología, estimación, recursos

Probabilidad	Impacto				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy bajo	Alto	Bajo	Medio	0,28	Mitigación
Probabilidad final	Impacto final				Impacto total	Respuesta efectiva
	Presupuesto	Planificación	Alcance	Calidad		
Alta	Muy bajo	Alto	Bajo	Medio	0,39	Mitigación

Riesgos derivados: N/A

Riesgo residual: N/A

Plan de contingencia: Búsqueda de alternativas a la descarga mediante la API oficial, como el raspado web, implementación del mismo

Planificación temporal de las contingencias: El plan de contingencia planteado se ejecutaría durante la fase de recogida de datos

Presupuesto para contingencias

Ítem	Concepto	Asignación (€)
1	Reestructuración metodología de descarga	61,52
2	Implementación de nuevo código	307,60
3	Descarga de los nuevos datos	0
	TOTAL	369,12

Monitorización:

- Las peticiones tienen que hacer sí o sí al servidor dónde se aloja la publicación de la que se quieren recuperar las respuestas, pues de lo contrario no se pueden recuperar todas.
- Sólo se puede recuperar una publicación por petición, incurriendo en limitaciones de peticiones en los servidores. Se despliega el programa de descargas en varias máquinas (incluidas las de compañeros del máster), con varios hilos en cada una realizando peticiones a diferentes servidores, para paralelizar el trabajo lo máximo posible.

7.2.3.3 Riesgo 04

Nombre: Pérdida de Twitter como fuente de datos

Descripción: Dado el cambio de dirección de Twitter en octubre de 2022, es perfectamente factible que los datos de esta red social dejen de estar disponibles, ora por el cierre del propio sitio, ora por rediseños de los niveles de acceso a los investigadores.

Categorías: tecnología, dependencias del proyecto, recursos

Probabilidad	Impacto				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy bajo	Alto	Alto	Medio	0,28	Asunción
Probabilidad final	Impacto final				Impacto total	Respuesta efectiva
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy bajo	Alto	Alto	Medio	0,28	No afectó al proyecto

Riesgos derivados: N/A

Riesgo residual: N/A

Plan de contingencia: Se asume el riesgo, pues escapa al control de los investigadores.

Planificación temporal de las contingencias: El riesgo quedará anulado en el momento en que se descarguen todos los datos de Twitter.

Presupuesto para contingencias: N/A

Monitorización:

- En marzo se actualizaron los niveles de acceso generales de la API, aunque no se vio afectada la API Académica
- En julio se eliminó la API académica como tal, sustituyéndola por una nueva de pago, inaccesible con los recursos del proyecto. Afortunadamente, se había descargado toda la información.

7.2.3.4 Riesgo 05

Nombre: Conflicto con los términos de uso de las instancias de Mastodon

Descripción: Existen instancias de Mastodon cuyos administradores explícitamente prohíben explícitamente el uso de los datos de sus usuarios para ciencia de datos, raspado web y similares. Por tanto, podría entrarse en conflictos éticos y potencialmente legales.

Categorías: usuarios, dependencias del proyecto, recursos

Probabilidad	Impacto				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Alta	Muy bajo	Bajo	Medio	Bajo	0,21	Eliminación
Probabilidad final	Impacto final				Impacto total	Respuesta efectiva
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy bajo	Medio	Medio	Bajo	0,15	Eliminación

Riesgos derivados: N/A

Riesgo residual: N/A

Plan de contingencia: Se eliminará el riesgo incluyendo en la fase de limpieza de datos una revisión manual de las instancias de las que se han extraído datos, comprobando sus términos de uso y códigos de conducta para ver si alguno de los documentos tiene cláusulas que entren en conflicto, y los toots provenientes de esas instancias se eliminarán del *dataset*.

Planificación temporal de las contingencias: La tarea será llevada a cabo durante la fase de limpieza de datos.

Presupuesto para contingencias: No se establece presupuesto de contingencias pues ya se incluye en la planificación y presupuestos una tarea para esta cuestión.

Monitorización: Se realizó el proceso de filtrado sin incidentes, aunque el proceso resultó más laborioso de lo esperado.

7.2.3.5 Riesgo 07

Nombre: Los experimentos considerados a priori no arrojan resultados significativos

Descripción: Se está trabajando en un línea de investigación en la que apenas hay nada hecho, y por tanto es probable que los experimentos que se planteen inicialmente no arrojen los resultados que cabría esperar.

Categorías: planificación, personal, recursos

Probabilidad	Impacto				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Alto	Muy alto	Alto	Alto	0,45	Mitigación
Probabilidad final	Impacto final				Impacto total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Alta	Alto	Muy alto	Alto	Alto	0,63	Mitigación

Riesgos derivados: N/A

Riesgo residual: N/A

Plan de contingencia: En caso de que esto se produzca, será necesario establecer nuevos experimentos y llevarlos a cabo para tratar de obtener resultados que puedan ser de interés en una publicación.

Planificación temporal de las contingencias: Dicho replanteamiento será llevado a cabo tras la primera realización de experimentos.

Presupuesto para contingencias:

Ítem	Concepto	Asignación (€)
1	Reestructuración de experimentos	553,68
	TOTAL	553,68

Monitorización:

- Los experimentos planteados inicialmente, basados en análisis de sentimientos, no dieron resultados significativos
- Se plantearon nuevos experimentos, basados en la vectorialización de textos
- Se decidió volver a preprocesar el *dataset* para corregir errores que se habían cometido (ver bitácora en sección 7.20)
- Se realizaron los experimentos, esta vez con resultados satisfactorios

7.3 Cierre del proyecto

7.3.1 Planificación final

La planificación final (que puede encontrarse desglosada en el fichero `planificacion_final.mpp` adjunto) consta de 8 tareas principales, para cada una de las cuales se ofrece un tabla desglosando sus subtareas:

- **Tarea 01:** Planificación del proyecto (véanse las tareas en el cuadro 7.21)
- **Tarea 24:** Estudio del estado del arte (véanse las tareas en el cuadro 7.22)
- **Tarea 33:** Obtención de datos (véanse las tareas en el cuadro 7.23)
- **Tarea 41:** Procesamiento de los datos (véanse las tareas en el cuadro 7.24)
- **Tarea 48:** Segundo preprocesamiento (véanse las tareas en el cuadro 7.25)
- **Tarea 51:** Reestructuración de experimentos (véanse las tareas en el cuadro 7.26)
- **Tarea 58:** Elaboración del artículo (véanse las tareas en el cuadro 7.27)
- **Tarea 60:** Cierre del proyecto (véanse las tareas en el cuadro 7.28)

Finalmente, el proyecto terminó el 27 de octubre de 2023, con una duración total de 267,25 días (41,25 días más de lo planificado originalmente). Se estableció el mes de agosto como vacaciones.

La razón de este incremento en el tiempo del proyecto, además de la inclusión de dos nuevos grupos de tareas que no estaban contemplados originalmente, son la materialización de los riesgos 02 y 07 (véase el cuadro 7.10). El primero implicó un sustancial incremento en la tarea de descarga de los datos de Mastodon (véase la sección 7.2.3.2 para la respuesta adoptada) y el segundo (véase la sección 7.2.3.5 para la respuesta adoptada) implicó una necesidad de establecer nuevos experimentos y ejecutarlos. Además, se detectaron errores en el primer preprocesado de los datos (tarea 43), por lo que se decidió volver a preprocesar el *dataset* para subsanarlos.

7.3.2 Informe final de riesgos

En las hojas de riesgos, de las que se informa debidamente en la sección 7.2.3, se encuentra la monitorización de los riesgos y la respuesta que se dio en caso de que se materializasen.

7.3.3 Presupuesto final

El presupuesto final tiene diferencias significativas con respecto al inicial. La primera es la obvia dilatación de algunas de las tareas. Si bien la más significativa ha sido la de las descargas de los datos, al no implicar trabajo por parte del proyectante, sólo afecta a la planificación, pero no al presupuesto.

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.21: Desglose de las tareas de correspondientes a la Planificación del proyecto (tarea con identificador 01). Las tareas 8, 9, 10, 11, 12 y 13 son hijas de la tarea 8. Las tareas 15 y 16 son hijas de la tarea 14. Las tareas 18, 19, 20 y 21 son hijas de la tarea 17.

Id	Nombre	Antecesora	Duración	Comienzo
2	Identificación de interesados	N/A	4 horas	2023-03-01
3	Captación de requisitos	2	5 días	2023-03-01
4	Definición de alcance	3	2 días	2023-03-06
5	Elaboración del OBS	4	2 horas	2023-03-07
6	Elaboración del PBS	5	2 horas	2023-03-07
7	Elaboración de la EDT	6	3 horas	2023-03-08
8	Planificación temporal	N/A	5,75 días	2023-03-08
9	Definición de actividades	7	3 horas	2023-03-08
10	Secuenciación de actividades	9	2 horas	2023-03-09
11	Estimación de recursos de las actividades	10	2 horas	2023-03-09
12	Estimación de la duración de las actividades	11	2 días	2023-03-09
13	Desarrollo del cronograma	12	2 días	2023-03-10
14	Planificación económica	N/A	4 días	2023-03-13
15	Estimación de costes	8	2 días	2023-03-13
16	Elaboración de presupuestos	15	2 días	2023-03-15
17	Planificación de la gestión de riesgos	N/A	5,75 días	2023-03-16
18	Identificación de riesgos	14	2 días	2023-03-16
19	Análisis cualitativo	18	3 horas	2023-03-17
20	Análisis cuantitativo	19	4 horas	2023-03-17
21	Planificación de respuestas	20	2 días	2023-03-20
22	Elaboración del acta de proyecto	17	2 días	2023-03-21
23	Reunión de aprobación del proyecto	22	4 horas	2023-03-22

Cuadro 7.22: Desglose de las tareas de correspondientes al Estudio del estado del arte (tarea con identificador 24)

Id	Nombre	Antecesora	Duración	Comienzo
25	Reunión de preparación de estado del arte	1	4 horas	2023-03-23
26	Descripción de la red social	25	7 días	2023-03-23
27	Espiral del silencio	26	5 días	2023-03-29
28	Diferencias del discurso	27	5 días	2023-04-03
29	Dinámicas de viralidad	28	7 días	2023-04-06
30	Hostigamiento a periodistas	29	3,5 días	2023-04-12
31	Ética en investigaciones en redes sociales	30	5 días	2023-04-14
32	Reunión de valoración del estado del arte	31	4 horas	2023-04-19

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.23: Desglose de las tareas de correspondientes a la Obtención de datos (tarea con identificador 33)

Id	Nombre	Antecesora	Duración	Comienzo
34	Reunión de preparación de obtención de datos	24	2 días	2023-04-19
35	Identificación de periodistas	34	6 días	2023-04-21
36	Desarrollo de código de descarga de Twitter	35	12 días	2023-04-26
37	Descarga de datos Twitter	36	7 días	2023-05-02
38	Desarrollo de código de descarga de Mastodon	37	12 días	2023-05-08
39	Descarga de datos de Mastodon	38	35 días	2023-05-17
40	Reunión de valoración de la obtención de datos	39	1,75 días	2023-06-09

Cuadro 7.24: Desglose de las tareas de correspondientes al Procesamiento de datos (tarea con identificador 41)

Id	Nombre	Antecesora	Duración	Comienzo
42	Reunión de preparación de procesamiento de datos	33	4 horas	2023-06-12
43	Limpieza de los datos	42	22 días	2023-06-13
44	Anonimización de los datos	43	3 días	2023-06-23
45	Análisis de los datos	44	16 días	2023-06-27
46	Elaboración de informe de resultados	45	3 días	2023-07-07
47	Reunión de valoración del procesamiento de datos	46	4 horas	2023-07-10

Cuadro 7.25: Desglose de las tareas de correspondientes al Segundo preprocesamiento (tarea con identificador 48)

Id	Nombre	Antecesora	Duración	Comienzo
49	Solución de problemas encontrados	47	7 días	2023-07-11
50	Ejecución del preprocesado	49	20 días	2023-07-13

Cuadro 7.26: Desglose de las tareas de correspondientes a la Reestructuración de experimentos (tarea con identificador 51)

Id	Nombre	Antecesora	Duración	Comienzo
52	Reunión de reestructuración	50	1 día	2023-07-28
53	Estudio de nuevo enfoque	52	7 días	2023-07-31
54	Planificación de experimentos	53	2 días	2023-08-04
55	Codificación de los experimentos	54	7 días	2023-08-07
56	Ejecución de los experimentos	55	5 días	2023-09-08
57	Análisis de los resultados	56	3 días	2023-09-13

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.27: Desglose de las tareas de correspondientes a la Elaboración del artículo (tarea con identificador 58)

Id	Nombre	Antecesora	Duración	Comienzo
59	Reunión de preparación del artículo	51	1 día	2023-09-15
60	Redacción de la metodología	59	2 días	2023-09-15
61	Redacción de los resultados	60	6 días	2023-09-18
62	Redacción de las conclusiones	61	4 días	2023-09-21
63	Redacción del estado del arte	62	7 días	2023-09-25
64	Redacción de la introducción	63	4 días	2023-09-29
65	Redacción del resumen y título	64	4 días	2023-10-03
66	Reunión de valoración del artículo	65	1 día	2023-10-05
67	Corrección de errores detectados	66	5 días	2023-10-06
68	Envío del artículo	67	2 horas	2023-10-06

Cuadro 7.28: Desglose de las tareas de correspondientes al Cierre de proyecto (tarea con identificador 69). Las tareas 72, 73, 74 y 75 son hijas de la tarea 71.

Id	Nombre	Antecesora	Duración	Comienzo
70	Traducción y adaptación de la memoria	58	11 días	2023-10-06
71	Análisis de dirección y gestión	N/A	18 días	2023-10-13
72	Elaboración de la planificación final	70	5 días	2023-10-13
73	Elaboración de presupuesto final	72	4 días	2023-10-18
74	Elaboración de informe final de riesgos	73	7 días	2023-10-20
75	Elaboración de informe de lecciones aprendidas	74	2 días	2023-10-26
76	Envío de la memoria	71	1 hora	2023-10-27

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

La segunda ha sido la necesidad de incluir dos partidas nuevas, debido a la materialización del riesgo 7 (ver sección 7.2.3.5). Se corresponden con las dos nuevas fases planteadas en la planificación final (sección 7.3.1).

La tercera, hilando con la cuestión de los riesgos, ha sido la ejecución de dos de las reservas de contingencia, las correspondientes a la reestructuración de experimentos. En el presupuesto final, esos fondos se restan a los del resto del proyecto, pues se emplean en la mitigación de riesgos. La tercera reserva de contingencia planteada en la partida original (cuadro 7.19) se ha eliminado del presupuesto final.

Las partidas con las que cuenta el presupuesto final son:

- **Partida 1 - Planificación del proyecto** (véase cuadro 7.13)
- **Partida 2 - Estudio del estado del arte** (véase cuadro 7.14)
- **Partida 3 - Obtención de datos** (véase cuadro 7.15)
- **Partida 4 - Procesamiento de los datos** (véase cuadro 7.16)
- **Partida 5 - Segundo preprocesamiento** (véase cuadro 7.16)
- **Partida 6 - Reestructuración de experimentos** (véase cuadro 7.16)
- **Partida 7 - Elaboración del artículo** (véase cuadro 7.17)
- **Partida 8 - Cierre de proyecto** (véase cuadro 7.18)
- **Partida 9 - Otros** (véase cuadro 7.19)

El proyecto tuvo un coste final de 13.242,18 €, o lo que es lo mismo, se ha incurrido en un sobrecoste del 32,46 % con respecto a lo presupuestado originalmente. Puede verse el presupuesto final resumido en el cuadro

Cuadro 7.29: Presupuesto final resumido

Cod.	Partida	Total
01	Planificación del proyecto	1.737,94
02	Estudio del estado del arte	2.122,44
03	Obtención de datos	2.076,30
04	Procesamiento de los datos	2.829,92
05	Segundo preprocesamiento	430,64
06	Reestructuración de experimentos	1.538,00
07	Elaboración del artículo	1.814,84
08	Cierre de proyecto	1.799,46
09	Otros	-1.107,36
	TOTAL	13.242,18

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.30: Desglose de la partida 1

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total (€)
01		Identificación de interesados	4	horas	15,38		61,52
02		Captación de requisitos	20	horas	15,38		307,60
03		Definición de alcance	8	horas	15,38		123,04
04		Elaboración del OBS	2	horas	15,38		30,76
05		Elaboración del PBS	2	horas	15,38		30,76
06		Elaboración de la EDT	3	horas	15,38		46,14
07		Planificación temporal					353,74
	001	Definición de actividades	3	horas	15,38	46,14	
	002	Secuenciación de actividades	2	horas	15,38	30,76	
	003	Estimación de recursos de las actividades	2	horas	15,38	30,76	
	004	Estimación de la duración de las actividades	8	horas	15,38	123,04	
	005	Desarrollo del cronograma	8	horas	15,38	123,04	
08		Planificación económica					246,08
	001	Estimación de costes	8	horas	15,38	123,04	
	002	Elaboración de presupuestos	8	horas	15,38	123,04	
09		Planificación de la gestión de riesgos					353,74
	001	Identificación de riesgos	8	horas	15,38	123,04	
	002	Análisis cualitativo	3	horas	15,38	46,14	
	003	Análisis cuantitativo	4	horas	15,38	61,52	
	004	Planificación de respuestas	8	horas	15,38	123,04	
10		Elaboración del acta de proyecto	8	horas	15,38		123,04
11		Reunión de aprobación del proyecto	4	horas	15,38		61,52
						TOTAL	1.737,94

Cuadro 7.31: Desglose de la partida 2

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)	
01	Reunión de preparación de estado del arte	4	horas	15,38	61,52	
02	Descripción de la red social	28	horas	15,38	430,64	
03	Espiral del silencio	20	horas	15,38	307,60	
04	Diferencias del discurso	20	horas	15,38	307,60	
05	Dinámicas de viralidad	28	horas	15,38	430,64	
06	Hostigamiento a periodistas	14	horas	15,38	215,32	
07	Ética en investigaciones en redes sociales	20	horas	15,38	307,60	
08	Reunión de valoración del estado del arte	4	horas	15,38	61,52	
					TOTAL	2.122,44

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.32: Desglose de la partida 3

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación de obtención de datos	8	horas	15,38	123,04
02	Identificación de periodistas	24	horas	15,38	369,12
03	Desarrollo de código de descarga de Twitter	48	horas	15,38	738,24
04	Descarga de datos Twitter	28	horas	0,00	0,00
05	Desarrollo de código de descarga de Mastodon	48	horas	15,38	738,24
06	Descarga de datos de Mastodon	140	horas	0,00	0,00
07	Reunión de valoración de la obtención de datos	7	horas	15,38	107,66
TOTAL					2.076,30

Cuadro 7.33: Desglose de la partida 4

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación de procesamiento de datos	4	horas	15,38	61,52 €
02	Limpieza de los datos	88	horas	15,38	1.353,44
03	Anonimización de los datos	12	horas	15,38	184,56
04	Análisis de los datos	64	horas	15,38	984,32
05	Elaboración de informe de resultados	12	horas	15,38	184,56
06	Reunión de valoración del procesamiento de datos	4	horas	15,38	61,52
TOTAL					2.829,92

Cuadro 7.34: Desglose de la partida 5

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Solución de problemas encontrados	28	horas	15,38	430,64
02	Ejecución del preprocesado	80	horas	0,00	0,00
TOTAL					430,64

Cuadro 7.35: Desglose de la partida 6

I 1	Descripción	Cantidad	Unidades	Precio	Total
01	Reunión de reestructuración	4	horas	15,38	61,52
02	Estudio de nuevo enfoque	28	horas	15,38	430,64
03	Planificación de experimentos	8	horas	15,38	123,04
04	Codificación de los experimentos	28	horas	15,38	430,64
05	Ejecución de los experimentos	20	horas	15,38	307,60
06	Análisis de resultados	12	horas	15,38	184,56
TOTAL					1.538,00

CAPÍTULO 7. DIRECCIÓN Y GESTIÓN DEL PROYECTO

Cuadro 7.36: Desglose de la partida 7

I 1	Descripción	Cantidad	Unidades	Precio (€)	Total (€)
01	Reunión de preparación del artículo	4	horas	15,38	61,52
02	Redacción de la metodología	8	horas	15,38	123,04
03	Redacción de los resultados	24	horas	15,38	369,12
04	Redacción de las conclusiones	16	horas	15,38	246,08
05	Redacción del estado del arte	28	horas	15,38	430,64
06	Redacción de la introducción	16	horas	15,38	246,08
07	Redacción del resumen y título	16	horas	15,38	246,08
08	Reunión de valoración del artículo	4	horas	15,38	61,52
09	Corrección de errores detectados	20	horas	15,38	307,60
10	Envío del artículo	2	horas	15,38	30,76
TOTAL					2.122,44

Cuadro 7.37: Desglose de la partida 8

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total (€)
01		Traducción y adaptación de la memoria	44	horas	15,38		676,72
02		Análisis de dirección y gestión					1.107,36
	001	Elaboración de la planificación final	20	horas	15,38	307,60	
	002	Elaboración de presupuesto final	16	horas	15,38	246,08	
	003	Elaboración del informe de riesgos	28	horas	15,38	430,64	
	004	Elaboración del informe de lecciones aprendidas	8	horas	15,38	123,04	
03		Envío de la memoria	1	horas	15,38		15,38
TOTAL							1.799,46

Cuadro 7.38: Desglose de la partida 9

I 1	I 2	Descripción	Cantidad	Unidades	Precio (€)	Subtotal (€)	Total (€)
01		Reservas de contingencia					1.107,36
	001	Reestructuración de experimentos 1	36	horas	15,38 €	553,68	
	002	Reestructuración de experimentos 2	36	horas	15,38 €	553,68	
TOTAL							1.107,36

7.3.4 Informe de lecciones aprendidas

Este proyecto ha tenido numerosos problemas ya desde su inicio, pues se partía de una idea bastante “nevulosa” de lo que se pretendía hacer, pero no había una línea de investigación firmemente marcada.

Esto en parte explica que hubiera que volver a replantear todos los experimentos y los problemas encontrados con la API. Además, en todo momento se contaba con la posibilidad de la pérdida de Twitter como fuente de datos, debido a los vaivenes de la compañía. El mismo contexto que daba valor a la investigación también la ponía en peligro.

De aquí podemos arrojar principalmente dos lecciones:

- Se ha cometido el error (demasiado) común de no tener un alcance y objetivos bien definidos a comienzo del proyecto, lo cual es más grave si tenemos en cuenta que no se puede echar la culpa a ningún cliente, pues no lo hay.
- Esto tiene que servir para poner en verdadero valor la gestión de riesgos, pues muchas de estas cuestiones estaban consideradas desde el principio y se han terminado manifestando.

Además, la cuestión de si continuar con esta línea de investigación o no está ahora mismo en el aire. Por un lado están las dificultades para la recogida de datos, pues como ya se ha dicho, el enfoque adoptado en este proyecto no es escalable de forma alguna. Por otro lado, hay que tener en cuenta la cuestión de las expectativas de privacidad de los usuarios del Fediverso. Con las medidas planteadas, es posible que este proyecto no cause revuelo, pero se navegan aguas turbulentas, y siempre se debe tener en cuenta que los datos son, en última instancia, personas. Y por último, la dificultad para obtener resultados relevantes del *dataset* podría ser indicativo de que el enfoque adoptado no es el correcto.

EL que se siga ahondando en esta línea de investigación o no está en gran medida sujeto a la recepción que el artículo tenga tanto en el mundo académico como entre la comunidad del Fediverso.

Bibliografía

- [AJNP20] Abi-Jaoude, Elia; Naylor, Karlina Treurnicht; Pignatiello, Antonio: Smartphones, social media use and youth mental health. *Cmaj*, 192(6):E136–E141, 2020.
- [Ak22] Al-khateeb, Samer: Dapping into the Fediverse: Analyzing What's Trending on Mastodon Social. In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, Springer International Publishing, pp. 101–110, 2022.
- [BB17] Bashir, Hilal; Bhat, Shabir Ahmad: Effects of social media on mental health: A review. *International Journal of Indian Psychology*, 4(3):125–131, 2017.
- [BLM22] Braghieri, Luca; Levy, Ro'ee; Makarin, Alexey: Social media and mental health. *American Economic Review*, 112(11):3660–3693, 2022.
- [BM19] Brüger, Niels; Milligan, Ian, eds. *The SAGE Handbook of Web History*. SAGE Publications Ltd, 2019.
- [BP22] Bajak, Aleszu; Padilla, Ramon: 'Hope' is out, 'fight' is in: Does tweeting divide Congress, or simply echo its divisions? *USA TODAY*, 2022.
- [De05] Dean, Jodi: Communicative capitalism: Circulation and the foreclosure of politics. *Cultural Politics*, 1(1):51–74, 2005.
- [fr20] Franzke, Aline Shakti; Bechmann, Anja; Zimmer, Michael; Ess, Charles M.: *Internet Research: Ethical Guidelines 3.0*, 2020.
- [Ga18] Garimella, Kiran; De Francisci Morales, Gianmarco; Gionis, Aristides; Mathioudakis, Michael: Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In: *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, pp. 913–922, 2018.
- [GJ17] Gupta, Itisha; Joshi, Nisheeth: Tweet normalization: A knowledge based approach. In: *2017 International Conference on Infocom Tech-*

BIBLIOGRAFÍA

- nologies and Unmanned Systems (Trends and Future Directions)(IC-TUS). IEEE, IEEE, pp. 157–162, 2017.
- [GM21] Garcia-Menendez, Angel: , Research project: Retrospective on decentralized/federated social media and analysis of its weak presence, 2021.
- [GSV23] Gilbert, Sarah; Shilton, Katie; Vitak, Jessica: When research is the context: Cross-platform user expectations for social media data reuse. *Big Data & Society*, 10(1):20539517231164108, 2023.
- [Gu21] Guntuku, Sharath Chandra; Purtle, Jonathan; Meisel, Zachary F; Merchant, Raina M; Agarwal, Anish: Partisan differences in twitter language among US legislators during the COVID-19 pandemic: Cross-sectional study. *Journal of medical Internet research*, 23(6):e27300, 2021.
- [Hi23] Hickey, Daniel; Schmitz, Matheus; Fessler, Daniel; Smaldino, Paul E; Muric, Goran; Burghardt, Keith: Auditing elon musk’s impact on hate speech and bots. In: *Proceedings of the International AAAI Conference on Web and Social Media*. volume 17, pp. 1133–1137, 2023.
- [HLJ16a] Hamilton, William L; Leskovec, Jure; Jurafsky, Dan: Cultural shift or linguistic drift? comparing two computational measures of semantic change. In: *Proceedings of the conference on empirical methods in natural language processing*. Conference on empirical methods in natural language processing. volume 2016. NIH Public Access, p. 2116, 2016.
- [HLJ16b] Hamilton, William L; Leskovec, Jure; Jurafsky, Dan: Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*, 2016.
- [Ho23] Holton, Avery E; Bélair-Gagnon, Valérie; Bossio, Diana; Molyneux, Logan: “Not their fault, but their problem”: Organizational responses to the online harassment of journalists. *Journalism Practice*, 17(4):859–874, 2023.
- [In13] Institute, Project Management: *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. Project Management Institute, 2013.
- [Je23] Jeong, Ujun; Sheth, Paras; Tahir, Anique; Alatawi, Faisal; Bernard, H Russell; Liu, Huan: Exploring platform migration patterns between twitter and mastodon: A user behavior study. *arXiv preprint arXiv:2305.09196*, 2023.
- [LCAT23] La Cava, Lucio; Aiello, Luca Maria; Tagarelli, Andrea: Get Out of the Nest! Drivers of Social Influence in the# TwitterMigration to Mastodon. *arXiv preprint arXiv:2305.19056*, 2023.

BIBLIOGRAFÍA

- [LQ13] Lin, Han; Qiu, Lin: Two sites, two voices: Linguistic differences between Facebook status updates and tweets. In: Cross-Cultural Design. Cultural Differences in Everyday Life: 5th International Conference, CCD 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part II 5. Springer, Springer Berlin Heidelberg, pp. 432–440, 2013.
- [Mi13] Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [SBVB22] Simchon, Almog; Brady, William J; Van Bavel, Jay J: Troll and divide: the language of online polarization. PNAS nexus, 1(1):pgac019, 2022.
- [Si01] Singhal, Amit et al.: Modern information retrieval: A brief overview. IEEE Data Eng. Bull., 24(4):35–43, 2001.
- [SP15] Sylwester, Karolina; Purver, Matthew: Twitter language use reflects psychological differences between democrats and republicans. PLoS one, 10(9):e0137422, 2015.
- [Uw23] Uwalaka, Temple; Amadi, Azubuike Fred; Nwala, Bigman; Wokoro, Peter: Online harassment of journalists in Nigeria: audience motivations and solutions. Media International Australia, 0(0):1329878X231206840, 2023.
- [Va23] Valero, Myriam Vidal: Thousands of scientists are cutting back on Twitter. Nature, 620:482–4, 2023.
- [W3] OStatus Community Group Wiki.
- [Wa14] Wang, Wenbo; Chen, Lu; Thirunarayan, Krishnaprasad; Sheth, Amit P: Cursing in english on twitter. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. Association for Computing Machinery, pp. 415–425, 2014.
- [Wa20] Waisbord, Silvio: Mob censorship: Online harassment of US journalists in times of digital hate and populism. Digital Journalism, 8(8):1030–1046, 2020.
- [Wa22] Waisbord, Silvio: Trolling journalists and the risks of digital publicity. Journalism Practice, 16(5):984–1000, 2022.
- [WHS21] Wells, Georgia; Horwitz, Jeff; Seetharaman, Deepa: Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show. The Wall Street Journal, 2021.
- [WT18] Webber, Christopher; Tallon, Jessica: ActivityPub. W3C recommendation, W3C, January 2018. <https://www.w3.org/TR/2018/REC-activitypub-20180123/>.

BIBLIOGRAFÍA

- [Ye23] Yeung, Ching-man Au; Liccardi, Ilaria; Lu, Kanghao; Seneviratne, Oshani; Berners-Lee, Tim: Decentralization: The Future of Online Social Networking. In: Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web. Association for Computing Machinery, New York, NY, USA, p. 187-199, 1 edition, 2023.
- [Zu19] Zuboff, Shoshana: The Age of Surveillance Capitalism. Profile Books, 2019.

Apéndices

Apéndice A

Plan de gestión de riesgos

A.1 Metodología

Para la gestión de riesgos del presente proyecto, se hará uso de la metodología que se expone en [In13], que se compone de los siguientes pasos:

1. Planificación de la gestión de riesgos
2. Identificación de los riesgos
3. Realización del análisis cuantitativo de los riesgos
4. Realización del análisis cuantitativo de los riesgos
5. Planificación de la respuesta de los riesgos
6. Monitorización y control de los riesgos

A.2 Roles y responsabilidades

Dada la escala del proyecto, se establecen dos roles que en última instancia recaen sobre la misma persona, el proyectante:

- Director del proyecto: encargado de la elaboración del plan de riesgos, así como del seguimiento.
- Responsable de riesgos: encargado de gestionar los riesgos, así como de actuar sobre los efectos de los mismos.

A.3 Presupuesto

En el cuadro A.3 se muestra el presupuesto asignado a la gestión de riesgos.

APÉNDICE A. PLAN DE GESTIÓN DE RIESGOS

Cuadro A.1: Presupuesto de la gestión de riesgos

Ítem	Concepto	Asignación (€)
1	Planificación de la gestión de riesgos	123,04
2	Identificación de los riesgos	46,14
3	Análisis de los riesgos	61,52
4	Planificación de respuesta de los riesgos	123,04
	Total	353,74

A.4 Calendario

Se muestra la calendarización de la gestión de riesgos en el cuadro A.4.

Cuadro A.2: Calendario de la gestión de riesgos

Actividad	Fecha
2023/03/16	Identificación de los riesgos
2023/03/17	Análisis cualitativo de los riesgos
2023/03/17	Análisis cuantitativo de los riesgos
2023/03/20	Planificación de respuesta de los riesgos

A.5 Categorías de riesgo

Se establecen las siguientes categorías de riesgos, a fin de permitir una mejor identificación, clasificación y gestión de los mismos. Cada riesgo puede estar dentro de una o más categorías.

1. Técnico
 - 1.1. Requisitos
 - 1.2. Tecnología
 - 1.3. Calidad
2. Organizacional
 - 2.1. Dependencias del proyecto
 - 2.2. Recursos
 - 2.3. Financiación
 - 2.4. Personal
3. Gestión de proyecto
 - 3.1. Estimación
 - 3.2. Planificación

- 3.3. Control
- 3.4. Comunicación
- 4. Externo
 - 4.1. Proveedores
 - 4.2. Tiempo
 - 4.3. Usuarios

A.6 Definiciones de probabilidad

En el cuadro A.6 podemos ver las definiciones de los niveles de probabilidad, así como el porcentaje equivalente y la descripción de cada nivel.

Cuadro A.3: Niveles de probabilidad, porcentajes equivalentes y sus descripciones

Nombre	Probabilidad	Descripción
Muy alta	90 %	Es muy probable que se materialice el riesgo
Alta	70 %	Es muy probable que se materialice el riesgo
Media	50 %	Existe cierta probabilidad de que el riesgo se materialice
Baja	30 %	Es poco probable que se materialice el riesgo
Muy baja	10 %	Es muy poco probable que se materialice el riesgo

A.7 Definiciones de impacto por objetivos

Las definiciones de los niveles de impacto sobre los diversos objetivos del proyecto, así como su escala numérica, pueden verse en el cuadro A.7.

A.8 Matriz de probabilidad de impacto

En el cuadro A.8 podemos ver la matriz de probabilidad de impacto.

A.9 Tolerancias

Dada la naturaleza del proyecto, se establece que la umbral de riesgo, esto es, la cota a partir de la cual se estima que el riesgo supone una verdadera amenaza para el proyecto, en 0,17. Los riesgos que no alcancen esa cota, *a priori*, no serán considerados como relevantes.

APÉNDICE A. PLAN DE GESTIÓN DE RIESGOS

Cuadro A.4: Niveles de impacto sobre los objetivos principales del proyecto, con sus respectivas definiciones y escala numérica.

Objetivos	Inapreciable (0,05)	Bajo (0,15)	Moderado (0,3)	Alto (0,55)	Crítico (0,9)
Coste	Incremento insignificante	Incremento de menos del 10 %	Incremento de más del 10 %	Incremento de más del 40 %	Incremento de más del 40 %
Tiempo	Incremento insignificantes	Incremento de menos del 5 %	Incremento de más del 5 %	Incremento de más del 10 %	Incremento de más del 20 %
Alcance	Reducciones inapreciables	Afectadas áreas poco importantes	Afectadas áreas importantes	Reducciones inaceptables	El resultado del proyecto es inútil
Calidad	Degradación inapreciable	Reducción poco importante	Importante reducción	Reducciones inaceptables	El resultado del proyecto es inútil

Cuadro A.5: Matriz de probabilidad de impacto

Probabilidad	Muy alta	0,90	0,05	0,14	0,27	0,50	0,81
	Alta	0,70	0,04	0,11	0,21	0,39	0,63
	Media	0,5	0,03	0,08	0,15	0,28	0,45
	Baja	0,30	0,02	0,05	0,09	0,17	0,27
	Muy baja	0,10	0,01	0,02	0,03	0,06	0,09
			0,05	0,15	0,30	0,55	0,9
			Inapreciable	Medio	Medio	Alto	Crítico
			Impacto				

A.10 Formatos de la documentación

No se establece ningún formato en específico para la documentación de los riesgos en este proyecto.

A.11 Seguimiento

El seguimiento se hará en reuniones periódicas semanales con el tutor, donde se valorarán los riesgos ya detectados, así como aquellos que pudieran ir surgiendo durante el proyecto. En caso de ser necesario tomar medidas o ejecutar los planes de contingencia, se informará a los interesados del proyecto.

Apéndice B

Cuestiones éticas

Como ya se ha expuesto hasta cierto punto en el capítulo 4, existen unas altas expectativas de privacidad [GSV23] por parte de los usuarios del Fediverso. El propio proyectante, como usuario del Fediverso, es partícipe de un entorno dónde tiene gran peso la cultura *hacker* y el rechazo a la vigilancia expuesta en el capítulo 3. Y no sería la primera vez que una investigación sobre el Fediverso termina en polémica, o incluso en un petición de retirada del artículo¹.

Para tratar de que el proyecto estuviera dentro de los márgenes esperables en lo que respecta a la ética en la investigación, se recurrió a la tercera edición de las Guías de Ética de la Association of Internet Researchers [fr20].

En el propio documento se reconoce lo problemático de la cuestión del consentimiento informado en los proyectos de Big Data, como podría considerarse este. Si bien se está haciendo uso de información “pública”, o bien disponible a través de una API que aporta información de usuarios que han aceptado unos determinados términos y condiciones, la mayoría de usuarios, si se les preguntase, no querrían que sus datos se utilizaran en experimentos.

No obstante, en redes sociales privativas como Twitter los propios usuarios han terminado por aceptar que esta práctica es común, y por ende no sería un caso tan problemático (al menos superficialmente). El problema se manifiesta cuando se pretende hacer uso de una plataforma como Mastodon, dónde los usuarios están mucho más preocupados por cuestiones de privacidad, habiendo muchos migrado al Fediverso huyendo precisamente de la vigilancia masiva (ver capítulo 3).

De acuerdo a la guía de la AoIR, aunque conseguir consentimiento informado de todos los usuarios resultaría impracticable, sí que se recomienda tomar medidas de mitigación.

¹Fuente: <https://www.sunclipse.org/wp-content/downloads/2020/01/open-letter.html>

En primer lugar, el listado de periodistas no fue escogido al azar. Uno de los principales problemas al recabar datos de los usuarios es que la información que estos cuelgan en Internet no es para que destaque. La idea es que estén diluidos entre la masa, y destacarlos haciéndolos partícipes de la creación de un *dataset* sería problemático. Empero, estos periodistas se apuntaron a esa lista precisamente porque buscaban ser conocidos por la comunidad. Estaban mostrando su voluntad de ser descubiertos por cualquiera abierta y voluntariamente. Es esta la razón principal de escoger ese listado, y no otro.

En lo que respecta a los usuarios cuyas interacciones se recabaron, hay que diferenciar entre las expectativas de privacidad del propio usuario y el de la instancia en que se aloja.

Los propios usuarios pueden marcar sus publicaciones con diferentes niveles de visibilidad, desde visible para cualquiera hasta únicamente para los que participan en la conversación. Además, en la configuración de perfil puede establecerse que esa persona no desea que los *bots* interactúen con él o ella. Finalmente, algunos incluso llegan a incluir etiquetas como `#nobot` en sus perfiles para indicar que no quieren que se recaben sus datos.

En el caso de las instancias, hay algunas que en sus términos y condiciones (o código de conducta) especifican que los datos de sus usuarios no pueden usarse para ciencia de datos o similares.

Para los usuarios se tomaron medidas automáticas. En primer lugar, al usar los *endpoints* públicos de la API de Mastodon, sólo se recuperaban las publicaciones visibles para cualquiera. Luego, una vez descargadas, se comprobaban los perfiles de los usuarios que habían hecho la publicación para ver si habían indicado alguna de las restricciones indicadas. En caso afirmativo, se descartaba el toot.

Para las instancias el proceso hubo de ser manual. Una vez descargados los toots se extrajeron las URL de todas las instancias de las que se había descargado al menos una publicación. Una vez obtenida esa lista, se revisaron manualmente los términos y condiciones y códigos de conducta de todas ellas, elaborando una lista negra de instancias. Una vez terminada, se eliminaron del *dataset* todas las publicaciones que provinieran de alguna de esas instancias.

Finalmente, se ha decidido no publicar en abierto el *dataset*, por las razones ya mencionadas. En el artículo figura información de contacto para que cualquier investigador interesado en hacer uso de él lo pida, y se considerará caso por caso si proveerlo o no.