



Universidad Nacional del Sur

TESIS DE DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN

*Herramientas para el análisis de grandes volúmenes de datos
en iniciativas de participación ciudadana*

Rocío Betsabé Hubert

BAHÍA BLANCA

ARGENTINA

2021

Prefacio

Esta Tesis se presenta como parte de los requisitos para optar al grado Académico de Doctorado en Ciencias de la Computación, de la Universidad Nacional del Sur y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otra. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el ámbito del Departamento de Ciencias e Ingeniería de la Computación durante el período comprendido entre el 1 de abril de 2017 y el 6 de diciembre de 2021, bajo la dirección del Dra. Elsa Estevez, Profesor Asociado del Departamento de Ciencias e Ingeniería de la Computación y del Dra. Ana Gabriela Maguitman, Profesor Asociado del Departamento de Ciencias e Ingeniería de la Computación.

.....

Rocío Betsabé Hubert

`rocio.hubert@cs.uns.edu.ar`

Departamento de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

Bahía Blanca, 6 de diciembre de 2021



UNIVERSIDAD NACIONAL DEL SUR
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el .../.../..., mereciendo la calificación de(.....)

Resumen

La adopción de nuevas tecnologías por parte de la sociedad, como el acceso a banda ancha y el uso de celulares, brinda nuevos canales para que los ciudadanos puedan expresar sus opiniones y conocimientos sobre distintas cuestiones de interés para la sociedad y los gobiernos. Estos nuevos canales no se limitan a las redes sociales, sino que incluyen también otros espacios como la sección de comentarios en periódicos digitales. Toda esta información resulta muy valiosa en los procesos de toma de decisiones de los gobiernos, ya que pueden utilizarla en diversas cuestiones de interés, que pueden variar desde el control y mantenimiento de servicios públicos al monitoreo de la opinión ciudadana sobre temas de políticas públicas. Las instituciones públicas tienen herramientas que permiten al ciudadano reportar problemáticas o realizar peticiones; sin embargo, no poseen soluciones tecnológicas que permitan buscar información relevante, procesarla y visualizarla eficientemente. En particular, existe una carencia de herramientas adecuadas que busquen y consoliden información distribuida en varias fuentes, que consideren la semántica de las opiniones ciudadanas, realizando procesamiento de texto y análisis de sentimiento.

En esta tesis se presentan tres herramientas. Una para el análisis visual de peticiones, quejas y reclamos ciudadanos realizados en un municipio, utilizando algoritmos de agrupaciones y diferentes técnicas de visualización. Otra herramienta para la recopilación, visualización y análisis de interacciones gobierno-ciudadanos en Twitter, realizando análisis de sentimiento y ofreciendo diferentes técnicas de visualización. La tercera herramienta es para la recuperación de información e identificación de ítems de información relevantes para el usuario, obteniendo información de Google Noticias, Reddit y Twitter, y mejorando los resultados utilizando un ciclo de aprendizaje activo.

Las contribuciones de esta tesis incluyen 1) un resumen de las iniciativas de participación ciudadana reflejadas en trabajos científicos, junto a un análisis y método de clasificación de las mismas; 2) el desarrollo de una plataforma para el análisis de peti-

ciones, quejas y reclamos ciudadanos; 3) el desarrollo de una herramienta para recuperar datos de redes sociales y analizar la actividad de cuentas de gobierno y la interacción de las mismas con los ciudadanos; y 4) el desarrollo de una plataforma para la recuperación de información de diferentes fuentes y la identificación de aquellos documentos más relevantes para el usuario. El desarrollo de estas herramientas y su validación mediante casos de estudio relevantes al área, tuvo como objetivo la identificación de técnicas de minería de datos, de análisis de sentimientos y de visualización adecuadas para la aplicación en el área de participación ciudadana.

Abstract

The adoption of new technologies by society, such as broadband access and smartphones, provides new channels for citizens to express their opinions and share their knowledge on various topics of interest to society and governments. These new channels include, but are not limited to, social networks, blogs and the comments section in digital newspapers. All this information is highly valuable in the government's decision-making processes since it can be used in various matters of interest, which can vary from the control and maintenance of public services to the monitoring of citizen opinion on public policy issues. Public institutions have tools that allow citizens to report problems or make requests; however, they do not have technological solutions that allow them to search for relevant information, process it and visualize it efficiently. In particular, there is a lack of adequate tools to search and consolidate information distributed in various sources that consider the semantic of citizens' opinions, performing text processing and sentiment analysis.

Three tools are presented in this thesis. One of these tools incorporates the visual analysis of citizen requests, complaints and claims made in a municipality, using clustering algorithms and different visualization techniques. Another tool provides mechanisms for the collection, visualization and analysis of government-citizen interactions on Twitter, performing sentiment analysis and offering different visualization techniques. The third tool was developed for information retrieval and identification of relevant documents for the user. The information is recovered from Google News, Reddit and Twitter, and the results are improved by an active learning cycle.

The contributions of this thesis include 1) a summary of the citizen participation initiatives reflected in scientific works, together with their analysis and classification methods; 2) the development of a platform for the analysis of citizen requests, complaints and claims; 3) the development of a tool with the capability of retrieving data from social networks and analyzing the activity of government accounts and their interaction with

citizens; and 4) the development of a platform for the retrieval of information from different sources and the identification of the most relevant documents for the user. The development of these tools and their validation through case studies relevant to the area, were aimed at identifying techniques for data mining, sentiment analysis and visualization suitable for its application in the area of citizen participation.

Índice general

| | |
|--|-----------|
| 1. Introducción | 1 |
| 1.1. Motivación | 1 |
| 1.2. Problema de investigación | 2 |
| 1.3. Solución propuesta | 3 |
| 1.4. Contribuciones | 4 |
| 1.5. Organización de la tesis | 5 |
| 2. Conceptos | 7 |
| 2.1. Participación electrónica | 7 |
| 2.2. Fuentes de información | 11 |
| 2.3. Técnicas informáticas | 16 |
| 2.4. Resumen | 30 |
| 3. Trabajos Relacionados | 31 |
| 3.1. Participación electrónica | 31 |
| 3.2. Técnicas informáticas en iniciativas de participación electrónica | 34 |
| 3.3. Uso de tecnologías en iniciativas de participación ciudadana | 41 |
| 3.4. Tipo de validaciones | 42 |
| 3.5. Evaluación del estado del arte | 45 |
| 3.6. Resumen | 51 |

| | |
|---|------------|
| 4. Herramientas informáticas para el análisis de comunicaciones ciudadanas | 53 |
| 4.1. Motivación | 53 |
| 4.2. Metodología de desarrollo | 54 |
| 4.3. Requerimientos | 55 |
| 4.4. Diseño | 59 |
| 4.5. Implementación y testeo | 62 |
| 4.6. Caso de estudio: Citymis | 64 |
| 4.7. Caso de estudio: Gestión colaborativa | 68 |
| 4.8. Discusión | 72 |
| 5. Herramientas informáticas para el análisis de interacción entre gobierno y ciudadanos | 75 |
| 5.1. Motivación | 75 |
| 5.2. Metodología de desarrollo | 76 |
| 5.3. Requerimientos | 77 |
| 5.4. Diseño | 82 |
| 5.5. Implementación y testeo | 86 |
| 5.6. Caso de estudio: Ministerios de países de Latinoamérica | 86 |
| 5.7. Caso de estudio: Secretarías de México | 97 |
| 5.8. Discusión | 108 |
| 6. Herramientas informáticas para análisis de comunicaciones públicas | 111 |
| 6.1. Motivación | 111 |
| 6.2. Metodología de desarrollo | 112 |
| 6.3. Requerimientos | 113 |
| 6.4. Diseño | 117 |
| 6.5. Implementación y testeo | 120 |

| | |
|---|------------|
| 6.6. Caso de estudio: Noticias de Google | 121 |
| 6.7. Caso de estudio: Noticias de Google | 124 |
| 6.8. Discusión | 125 |
| 7. Conclusiones y trabajo a futuro | 129 |
| 7.1. Este trabajo | 129 |
| 7.2. Discusión | 131 |
| 7.3. Trabajo futuro | 133 |
| 7.4. Publicaciones | 134 |
| A. Clasificación de iniciativas de participación ciudadana | 137 |

Índice de figuras

| | |
|---|----|
| 2.1. Ejemplo de <i>hierarchical greedy clustering</i> | 21 |
| 3.1. Línea de tiempo de la evolución en el uso de las tecnologías. | 41 |
| 3.2. Proporción del tipo de validación con respecto a las tecnologías utilizadas. | 44 |
| 3.3. Relación entre niveles y áreas de participación. | 45 |
| 3.4. Uso de técnicas en cada nivel y área de participación. | 47 |
| 3.5. Relación entre las áreas de participación. | 47 |
| 3.6. Número de iniciativas por origen en cada nivel y área de participación. | 49 |
| 3.7. Tecnologías en cada nivel de participación por las iniciativas dirigidas por el gobierno y por la investigación. | 50 |
| 4.1. Diagrama de caso de uso. | 55 |
| 4.2. Diagrama de componentes. | 60 |
| 4.3. Funcionalidad de navegación de datos estadísticos. | 65 |
| 4.4. Mapa de calor para el servicio «Higiene Urbana». | 67 |
| 4.5. Mapa de agrupaciones para el servicio «Alumbrado Público». | 68 |
| 4.6. Funcionalidad de navegación de datos estadísticos. | 70 |
| 4.7. Mapas de calor para la Ciudad de Buenos Aires. | 70 |
| 4.8. Mapa de agrupaciones para la Ciudad de Buenos Aires. | 71 |
| 5.1. Diagrama de caso de uso. | 78 |
| 5.2. Diagrama de componentes. | 83 |

| | | |
|-------|--|-----|
| 5.3. | Diagramas <i>event drop</i> : niveles de actividad de cada sector. | 88 |
| 5.4. | Nubes de palabras: temas tratados por el sector Desarrollo Social. | 90 |
| 5.5. | Diagramas de Euler: uso de recursos compartidos por el sector Educación. | 90 |
| 5.6. | Diagramas de Euler: uso de recursos compartidos por cada sector público. | 91 |
| 5.7. | Diagramas de Hasse Extendido: respuestas ciudadanas a los Ministerios/Secretarías de Trabajo de cada país. | 92 |
| 5.8. | Diagramas de Hasse Extendidos: respuestas ciudadanas a todos los Ministerios/Secretarías de cada país. | 93 |
| 5.9. | Gráficos de radar: emociones identificados en sector Salud. | 94 |
| 5.10. | Gráficos de líneas: correlaciones entre recursos multimedia y respuestas ciudadanas en el sector Salud en Argentina. | 95 |
| 5.11. | Gráficos de barras: correlaciones entre recursos multimedias y respuestas ciudadanas en Argentina. | 96 |
| 5.12. | Gráficos de serie de tiempo: niveles de actividad en México. | 99 |
| 5.13. | Nubes de palabras: temas tratados en México. | 100 |
| 5.14. | Diagramas de Euler: uso de recursos multimedia en México. | 101 |
| 5.15. | Diagramas de Hasse Extendidos: respuestas ciudadanas en México. | 102 |
| 5.16. | Diagramas de radar: emociones identificadas en México. | 103 |
| 5.17. | Gráfico de burbujas: emociones a lo largo del tiempo en el sector Salud. | 105 |
| 5.18. | Diagramas de Sankey: relación entre uso de recursos multimedia y respuestas ciudadanas en México. | 107 |
| 6.1. | Diagrama de caso de uso. | 113 |
| 6.2. | Diagrama de dispersión: comportamiento del modelo. | 116 |
| 6.3. | Diagrama de componentes. | 117 |
| 6.4. | Evolución del comportamiento del modelo <i>domestic_violence</i> | 123 |
| 6.5. | Evolución del comportamiento del modelo <i>criminalization_sex_work</i> | 125 |

Índice de tablas

| | |
|---|-----|
| 3.1. Principales aspectos de las iniciativas de e-Participación. | 33 |
| 4.1. Ejemplo de una entrada en la base de datos provista por Citymis. | 61 |
| 5.1. Ejemplos de palabras asociadas con las ocho emociones primarias. | 82 |
| 5.2. Los ministerios y las cuentas oficiales de Twitter bajo análisis. | 87 |
| A.1. Clasificación de las iniciativas identificadas en la revisión de literatura. . . | 138 |

Siglas

| | |
|---------------|--|
| API | Interfaz de Programación de Aplicaciones |
| DBSCAN | Density-based Spatial Clustering Of Applications With Noise |
| FODA | Análisis de Fortalezas, Oportunidad, Debilidades y Amenazas |
| GIS | Sistema de Información Geográfica |
| ODI | <i>Open Data Institute</i> o Instituto de Datos Abiertos |
| PI2 | Pregunta de Investigación 2 |
| PI1 | Pregunta de Investigación 1 |
| PQR | Peticiones, Quejas y Reclamos |
| SUACI | Sistema Único de Atención Ciudadana |
| TIC | Tecnologías de la Información y la Comunicación |
| UNDESA | Departamento de Asuntos Económicos y Sociales de las Naciones Unidas |
| UNS | Universidad Nacional del Sur |

Capítulo 1

Introducción

En este capítulo introductorio se presenta el contexto de la participación ciudadana, indicando la motivación para realizar este trabajo de investigación en base a oportunidades de desarrollo actuales en el área (Sección 1.1). A continuación, se indican las preguntas de investigación que guían esta tesis (Sección 1.2) y la consiguiente solución propuesta (Sección 1.3). Se listan luego las contribuciones que surgen del desarrollo de la tesis (Sección 1.4). Finalmente, en la Sección 1.5 se informa la organización del resto del documento.

1.1. Motivación

Los problemas que enfrentan las sociedades actuales son complejos y de diversas índoles, es por esto que los gobiernos han reconocido una falta de capacidad para poder dar respuesta o soluciones a esos problemas. Por lo tanto, cada vez es más necesaria la participación conjunta de distintos actores del ecosistema social, incluidos el gobierno, la academia, el sector productivo y también los ciudadanos, que trabajen colaborativamente en el desarrollo de las soluciones. Esto da origen a nuevas formas de gobernanza, donde comienza a tomar mucha importancia la participación ciudadana.

Las causas de participación ciudadana son diversas, pero es un hecho que muchos ciudadanos participan activamente en redes sociales, usan aplicaciones móviles y expresan sus opiniones abiertamente sobre distintas cuestiones de interés para la sociedad (por ejemplo, problemas de seguridad, presencia de baches en las calles, etc.) y para los gobiernos (por ejemplo, opinión sobre temas de políticas públicas o sobre el uso del presupuesto

de gobierno, entre otros) usando distintos tipos de plataformas digitales. Si bien la información publicada por los ciudadanos es de suma relevancia para los procesos de toma de decisiones de gobierno (RBM16), las instituciones públicas carecen de las capacidades humanas e institucionales para poder hacer un uso efectivo de la misma para innovar en la entrega de servicios públicos (BEJ16). En particular, no poseen soluciones tecnológicas que permitan buscar información relevante a un propósito en toda la marea de información provista por los ciudadanos, y que permitan procesar eficientemente y visualizar la información relevante. Los desafíos para proveer este tipo de herramientas radica en buscar y consolidar información distribuida en varias fuentes (como, por ejemplo, redes sociales y plataformas propietarias de gobierno); en cómo vincular y combinar datos u opiniones de ciudadanos considerando la semántica de los mismos; y en cómo utilizar esos datos en análisis predictivos para mejorar la calidad de vida de los ciudadanos o hacer un uso más efectivo de los recursos públicos (JMZ15), entre otros.

A fin de asegurar que la participación ciudadana sea efectiva, es decir que se entregan nuevos canales en los cuales se facilita la interacción entre ciudadanos y gobiernos, surgen líneas de investigación aplicadas al uso de herramientas digitales en el área de e-Participación. Por otro lado, en los últimos años han surgido nuevas tecnologías que tienen que ver con la recolección y el procesamiento de grandes volúmenes de datos. Esto hace necesaria la búsqueda de mecanismos para aplicar tecnologías que nacen de la ciencia de datos y de la inteligencia artificial para procesar toda esta información provista por ciudadanos en distintos canales, relacionados a sus necesidades y opiniones.

El análisis del estado del arte indica que hay pocas investigaciones en la intersección de áreas relacionadas con ciencia de datos, la ingeniería de software y la participación ciudadana. Esta tesis busca tratar de dar respuesta a esa necesidad, planteando el problema de investigación que se indica en la siguiente sección.

1.2. Problema de investigación

Dada la motivación presentada en la sección anterior, el problema de investigación definido en esta tesis se especifica en las siguientes preguntas.

Pregunta de Investigación 1 (PI1): ¿Cuáles son las técnicas de minería de datos y de visualización más convenientes para el análisis de opiniones, quejas y reclamos vertidos en diversas plataformas digitales en el contexto de participación electrónica?

Pregunta de Investigación 2 (PI2): ¿Qué componentes y técnicas pueden ser parte de una infraestructura de software que haga uso de información volcada en plataformas digitales para facilitar la toma de decisiones por parte de funcionarios públicos?

La PI1 se centra en la búsqueda y el análisis de diferentes plataformas de participación ciudadana, y particularmente en la identificación de las técnicas de minería de datos y de visualización utilizadas para llevar a cabo su fin. Esto se realiza con el objetivo de reconocer las técnicas adecuadas según la finalidad de la plataforma a desarrollar, dependiendo también del nivel y área de participación.

La PI2 busca la implementación de herramientas digitales con el fin de facilitar la toma de decisiones de funcionarios públicos, en base a datos recolectados de diversas fuentes digitales. Esta implementación incluye no solo la selección adecuada de técnicas de minería de datos y de visualización acorde al objetivo de la plataforma, sino también la conceptualización de los datos a utilizar, el diseño de la herramienta y finalmente su validación mediante casos de estudio relevantes al área de participación ciudadana.

1.3. Solución propuesta

En base a las preguntas de investigación y el análisis de iniciativas de participación ciudadana presentes en la literatura, se determinó que la participación ciudadana tiene lugar en diferentes plataformas que facilitan la interacción gobierno-ciudadano. Estas plataformas pueden ser propietarias, desarrolladas por los municipios para manejar los peticiones, quejas y reclamos (PQR), o pueden basarse en el análisis de opiniones expresadas por los ciudadanos en redes sociales o comentarios en noticias de periódicos digitales. Es por esto que se desarrollaron herramientas que buscan resolver problemáticas utilizando diferentes fuentes de información, a fin de clarificar el tipo de opiniones y datos encontrados.

Una de las herramientas propuestas en esta tesis se enfoca en PQR ciudadanos, en cómo representarlos para su posterior análisis y definición de la estrategia para su resolución. La segunda herramienta se enfoca en las interacciones gobierno-ciudadano en redes sociales, específicamente en Twitter. Se encarga tanto de recuperar información de interés como en ofrecer visualizaciones y posibles técnicas de minería de datos a utilizarse en el análisis. Finalmente, la tercera herramienta busca mejorar el conjunto final de datos para el análisis. Dada la cantidad de información disponible en medios digitales, se desarrolló

una herramienta para que el usuario defina la temática y el estilo de contenido que desea posteriormente analizar (posiblemente utilizando otra plataforma), valiéndose de técnicas de minería de datos.

1.4. Contribuciones

La primera contribución de esta tesis nace como resultado de la revisión de la literatura (ver Capítulo 3). Se ofrece un resumen de todas aquellas iniciativas de participación ciudadana reflejadas en trabajos científicos publicados que cuentan con alguna herramienta o plataforma desarrollada, clasificadas en base a un método propuesto por la tesista (ver Capítulo 3). Junto con esta recopilación, se identificaron también las tecnologías más utilizadas en los últimos años y la evolución de su uso en el área de participación electrónica a lo largo del tiempo (ver Capítulo 3). Se realiza además un análisis de las relaciones entre las diferentes categorías de la clasificación propuesta (Sección 3.5): nivel y área de participación, origen de la iniciativa, técnicas utilizadas y tipo de validaciones llevadas a cabo, haciendo foco en las tecnologías utilizadas.

Las siguientes contribuciones están asociadas al desarrollo de plataformas que ofrecen facilidades para la recuperación, el análisis y la visualización de datos en el contexto de iniciativas de e-Participación. La primera de estas contribuciones es el desarrollo de una plataforma para el análisis de PQR ciudadanos, donde se aplican los algoritmos de agrupamiento *Density-based spatial clustering of applications with noise* (DBSCAN) (EKS⁺96) y *hierarchical greedy clustering* para facilitar el análisis y visualización de PQR en un mapa de la ciudad (Capítulo 4). La segunda es la implementación de una herramienta para recuperar datos de redes sociales (en particular, de Twitter) y analizar la actividad de cuentas de gobierno y la interacción de las mismas con los ciudadanos (Capítulo 5). En este trabajo se destaca el análisis realizado sobre cuentas de países de habla hispana, el uso de análisis de sentimiento para identificar diferentes emociones en los comentarios ciudadanos (en base a la rueda de emociones de Plutchik), y la aplicación y desarrollo de diferentes visualizaciones que se adaptan al contexto particular de interacciones gobierno-ciudadano, como son la nube de palabras, el diagrama de Euler, el diagrama de Sankey o el diagrama de Hasse extendido (propuesto por la tesista). Finalmente, la última contribución es el desarrollo de una herramienta para la recuperación de información de diferentes fuentes de información (en particular, de Google Noticias, Reddit y Twitter) y la identi-

ficación de aquellos ítems de información más relevantes para el usuario (Capítulo 6). En este trabajo se destaca el uso de modelos de clasificación y la aplicación de aprendizaje activo para la rápida y eficaz identificación automática de noticias relevantes en base a un pequeño número de noticias clasificadas manualmente por el usuario.

1.5. Organización de la tesis

El resto de la tesis se encuentra organizada como se explica a continuación. En el Capítulo 2 se introducen los conceptos necesarios para el desarrollo de la tesis, en particular las nociones relacionadas a la participación electrónica, a las diferentes fuentes de información y a las técnicas informáticas para la recuperación, análisis y visualización de datos. Estos conceptos serán la base tanto para presentar los trabajos relacionados, como para el desarrollo de las herramientas propuestas. En el Capítulo 3 se presenta una revisión de la literatura para identificar los aspectos más relevantes de la e-Participación. También se realiza un análisis de trabajos relacionados que presentan una herramienta o plataforma de participación ciudadana, explorando las relaciones entre niveles y áreas, y el uso de tecnologías en diferentes niveles, áreas y origen de la iniciativa.

En el Capítulo 4 se presenta una plataforma para el agrupamiento y exploración visual de PQR ciudadanos. Esta herramienta presenta diferentes visualizaciones para observar la distribución de PQR a lo largo de la ciudad y en los diferentes servicios que ofrece un municipio. El Capítulo 5 propone una herramienta para analizar las interacciones entre el gobierno y los ciudadanos en Twitter aplicando técnicas de análisis de sentimientos y diferentes visualizaciones. Esta plataforma también permite analizar, por medio de algunas visualizaciones, los comentarios de ciudadanos en respuesta a noticias en periódicos digitales. La tercera plataforma se presenta en el Capítulo 6, destinada a la recuperación y el refinamiento de artículos de información relevantes para un usuario experto utilizando modelos de clasificación y ciclos de aprendizaje activo. El usuario puede elegir recuperar información de Google Noticias, Reddit y Twitter en un mismo modelo de clasificación, que reflejará alguna temática bajo análisis.

Finalmente, en el Capítulo 7 se explica el trabajo desarrollado en esta tesis y se discute a la luz de las preguntas de investigación presentadas en el Capítulo 2. A continuación, se especifican una serie de cuestiones que pueden abordarse en trabajos futuros. Adicional-

mente, en esta tesis se incluye el Apéndice A, que indica la clasificación de los trabajos identificados en la Sección 3.2.

Capítulo 2

Conceptos

En este capítulo se presentarán los conceptos transversales al contenido de esta tesis relacionados a la participación electrónica y las iniciativas de participación ciudadana (Sección 2.1), a las fuentes de datos que se pueden utilizar para estas iniciativas (Sección 2.2), y a las técnicas informáticas aplicables para la recuperación, análisis y representación de estos datos (Sección 2.3). Finalmente, en la Sección 2.4 se ofrece un resumen de lo abordado en este capítulo.

2.1. Participación electrónica

El término e-Participación (o participación electrónica) se originó a principios de la década de los 2000 refiriéndose al uso de las Tecnologías de la Información y la Comunicación (TIC) para ampliar y profundizar la participación política al permitir a los ciudadanos comunicarse entre ellos y con sus representantes electos (Mac04). Las TIC hicieron posible que diversos procesos (tales como administración, prestación de servicios, toma de decisiones y formulación de políticas públicas) involucren tanto al gobierno como a sus ciudadanos. Es por esto que el término e-Participación está fuertemente relacionado con términos tales como e-Gobierno, e-Gobernanza y e-Democracia.

La complejidad de los procesos de e-Participación resulta de la gran cantidad de áreas de participación, de las partes interesadas involucradas, de los niveles de participación y de las etapas en la formulación de políticas (FLL⁺06). Varios autores definieron diferentes

dimensiones para entender correctamente las iniciativas de e-Participación. Probablemente, las dimensiones más frecuentemente utilizadas son niveles y áreas de participación, y tecnologías utilizadas. Estas últimas dependen, dada su rápida evolución, del período de tiempo analizado, pero los niveles y áreas están extensamente documentados.

El nivel de e-Participación de una iniciativa está determinado por el tipo de interacción gobierno-ciudadano y el modo en que las contribuciones de los ciudadanos son utilizadas por el gobierno en los procesos de toma de decisiones y formulación de políticas. Algunos autores identifican tres niveles de e-Participación: *información*, *consulta* y *participación activa* (PL⁺01; Mac04). Otros, como la Asociación Internacional para la Participación Pública, proveen un modelo con cinco niveles de participación, incluyendo *información*, *consulta*, *compromiso*, *colaboración* y *empoderamiento* (iap20). Estos modelos no tienen en cuenta que gran parte del flujo de información es iniciado por los ciudadanos o por Organizaciones No Gubernamentales. En consecuencia, en (FLL⁺06) se propone un modelo de cuatro **niveles de e-Participación**, que es el utilizado en esta tesis:

- *e-Información* – se refiere a un canal unidireccional que provee información, ya sea del gobierno (mediante sitios web) o de los ciudadanos (como pueden ser las peticiones electrónicas).
- *e-Consulta* – está relacionado a un canal bidireccional limitado. Esto permite a las partes interesadas mantenerse informadas y contribuir con su opinión sobre asuntos específicos en las iniciativas oficiales.
- *e-Colaboración* – es un canal bidireccional mejorado. Aquí, las partes interesadas tienen un papel activo en la propuesta y formulación de políticas. Sin embargo, la responsabilidad de la decisión final recae en los funcionarios gubernamentales.
- *e-Empoderamiento* – representa el nivel más alto de participación, donde la ciudadanía es la que toma la decisión final. Por ejemplo, los presupuestos participativos.

Las TIC pueden ser utilizadas para facilitar la participación ciudadana en diferentes áreas. Estas áreas pueden ser tan diversas como los campos de aplicación y varían de acuerdo a cada iniciativa. Además, nuevas áreas son creadas y agregadas de acuerdo a las nuevas necesidades y las circunstancias políticas, sociales y culturales, entre otras. En particular, el proyecto DEMO-net (nC⁺06) usa las siguientes **áreas de e-Participación**:

- *Provisión de información* – para estructurar, representar y gestionar la información en contextos de participación. Un ejemplo es un sistema donde los ciudadanos se informan acerca de los últimos comunicados y eventos (BFF15).

- *Construcción comunitaria* – para apoyar que las personas formen comunidades a fin de avanzar en agendas compartidas, y para empoderar a dichas comunidades. Por ejemplo, la aplicación móvil *Täsä* permite a los ciudadanos realizar contribuciones sobre la ciudad que otros usuarios pueden comentar y votar (Thi16b).
- *Consulta* – para iniciativas que permiten a las partes interesadas contribuir con su opinión, de manera privada o pública, sobre cuestiones específicas. Un ejemplo es una plataforma que sirve como canal para que los ciudadanos expresen sus opiniones, solicitudes y quejas (OW09).
- *Campaña* – para iniciativas relacionadas con la protesta, el hacer *lobby*, la petición y otras formas de acción colectiva, a excepción de las campañas electorales. Un ejemplo es el proyecto *ZmapujTo* para mapear los vertederos ilegales realizado por investigadores y ciudadanos (KH14a).
- *Campaña electoral* – para apoyar a políticos, partidos políticos y cabilderos en el contexto de campañas electorales. Por ejemplo, el proyecto *DemokratiTorget* busca detener el declive en la participación de sus votantes y alentar a los electores a influir en los procesos de toma de decisiones políticas (SFS11).
- *Deliberación* – para apoyar discusiones virtuales, en grupos grandes o pequeños, permitiendo el debate crítico y la reflexión sobre temas específicos. Por ejemplo, la plataforma de deliberación presentada en (DCPS08) para superar los obstáculos a la participación de los procesos locales de Agenda 21 (Age21) en Italia.
- *Discurso* – para apoyar el análisis y la representación del discurso. Un ejemplo es la técnica webométrica propuesta en (JPWP15) que permite mapear el discurso público entre los participantes que mencionan la cuenta de Twitter de una agencia gubernamental en Nueva York y San Francisco.
- *Mediación* – para resolver disputas o conflictos en un contexto en línea. Por ejemplo, la plataforma *EngagementHQ*¹ tiene una funcionalidad que permite rastrear y moderar todos los comentarios publicados en la plataforma para garantizar que las discusiones se mantengan en un tono adecuado.
- *Planificación territorial* – relacionados al urbanismo y la evaluación ambiental. Por ejemplo, un proyecto para visualizar los cambios planificados en la ciudad de Viena, Austria (RBS⁺14).
- *Encuesta* – para medir la opinión pública y los sentimientos de los usuarios. Por ejemplo, la plataforma *MyUniversity*, usada por la Universidad de Ambato (Ecu-

¹ *EngagementHQ* - <https://www.bangthetable.com/engagementhq-community-software/>

dor), cuenta con un conjunto de herramientas que permiten debatir, votar y usar cuestionarios en línea (PL18).

- *Votación* – relacionados a la votación pública en elecciones, referendos o plebiscitos locales. Por ejemplo, el framework propuesto en (Mur10) que integra herramientas de conocimiento electrónico, consulta electrónica y voto electrónico que facilitan los procesos de toma de decisiones.

Las iniciativas de e-Participación son creadas en respuesta a diferentes necesidades, tal como puede ser los ciudadanos realizando pedidos a sus gobiernos, los gobiernos tratando de involucrar a los ciudadanos en diferentes tipos de decisiones (por ejemplo, decidir cómo utilizar el presupuesto gubernamental), o los investigadores explorando nuevas técnicas para la e-Participación (por ejemplo, cómo analizar los debates públicos). De este modo, se pueden clasificar el **origen de las iniciativas** en función de las siguientes categorías:

- Liderados por una sociedad civil: aquellas iniciadas o creadas por sociedades civiles o por los ciudadanos. Por ejemplo, la iniciativa de participación pública de los sindicatos de docentes en la plataforma *Liberopinion* (QG18).
- Dirigidas por el gobierno: aquellas iniciadas por agencias gubernamentales. Por ejemplo, la plataforma de discusión *XPIR*, desarrollada por el Ministerio de Educación para ayudar a la interacción entre los científicos y el gobierno (VKIS16).
- Lideradas por la investigación: aquellas que forman parte de un proyecto de investigación. En muchos casos, en estas iniciativas se utilizan datos gubernamentales abiertos. Por ejemplo, una aplicación que permite una comunicación estructurada entre los actores de la administración pública y los ciudadanos (BFR16).

En particular, se pueden agregar otras dos categorías de origen de iniciativas, dependiendo del nivel de participación de las partes. Estas categorías son:

- Dirigidos por el gobierno con participación de investigadores: se refiere a proyectos de investigación que implican modificaciones o mejoras propuestas por investigadores sobre iniciativas implementadas por gobiernos. Esta categoría es usada para indicar tanto la plataforma gubernamental como las actualizaciones propuestas por los investigadores. Por ejemplo, las visualizaciones de los resultados de la agrupación de texto realizadas por investigadores sobre los informes ciudadanos obtenidos de la herramienta de e-Participación desarrollada por el gobierno de Yakarta (DHS⁺17).
- Liderados por la investigación con ayuda gubernamental: se refiere a proyectos de investigación que utilizan datos no divulgados de iniciativas gubernamentales o que

proponen iniciativas que son utilizadas por los gobiernos. Por ejemplo, el estudio realizado en (POB+12) que tiene como objetivo identificar medidas de tránsito implementables a corto plazo propuestas por los ciudadanos a través de los canales digitales proporcionados por el gobierno.

Independientemente de las diversas dimensiones, y dada la naturaleza de la e-Participación, las TIC son un habilitador clave para promover la participación ciudadana. Con más de quince años de desarrollo, se han utilizado diversas herramientas y tecnologías en iniciativas y aplicaciones para incentivar el uso de las TIC a fin de promover la participación ciudadana y acercar la ciudadanía y el gobierno. En la literatura existen muchos artículos con referencias a diversas técnicas y mecanismos utilizados para abordar la e-Participación. Sin embargo, a medida que aparecen nuevas tecnologías y nuevos resultados de investigaciones, se expande el abanico de técnicas y mecanismos para utilizar y mejorar la participación ciudadana.

En secciones posteriores, se presentarán diferentes técnicas que se utilizan en iniciativas de participación ciudadana y en el desarrollo de esta tesis. En particular, en el Capítulo 3.5 se presenta una evaluación del estado del arte con una guía de aplicación de tecnologías dependiendo de nivel y área de la iniciativa. Así mismo, en el Apéndice A se clasifican los ejemplos mencionados en este capítulo según nivel, área, origen y tecnologías utilizadas.

2.2. Fuentes de información

Los datos que pueden utilizarse para desarrollar nuevas herramientas de e-Participación, ya sea analizando la interacción gobierno-ciudadano, la opinión de la ciudadanía acerca de algún tema en particular o simplemente recopilando información útil para mejorar su posterior análisis, son tan diversos como las iniciativas de participación ciudadana en sí mismas.

En el desarrollo de esta tesis, se utilizaron tres principales fuentes de información para el análisis y visualización de datos relacionados con la interacción gobierno-ciudadano: (1) datos abiertos, compartidos y cerrados relacionados a la generación de PQR ciudadanos; y datos obtenidos de (2) redes sociales y (3) periódicos digitales.

2.2.1. Datos abiertos, compartidos y cerrados

El concepto de datos abiertos es una práctica que busca la disponibilidad de determinados tipos de datos de forma libre, sin ningún tipo de restricción, ya sea monetaria o de otro tipo, que pueda ser usada y completamente procesada con al menos alguna herramienta de software gratis, libre o de código abierto. Se busca entonces que cualquiera pueda acceder, usar, modificar y compartir esta información, sujeto también a mantener ese estado libre (Ope20). En particular, en el contexto de gobernanza electrónica, la accesibilidad a la información gubernamental puede facilitar la transparencia, responsabilidad y participación ciudadana, además de apoyar y fomentar la innovación tecnológica y el crecimiento económico al tercerizar el desarrollo de nuevas aplicaciones y servicios digitales (Gra14; ope18).

Dadas estas características, los datos abiertos son claves para el desarrollo de ciudades inteligentes y para promover la participación ciudadana. Actualmente, muchos países y ciudades cuentan con portales digitales que permiten acceder a datos de gobierno. Estos datos pueden estar relacionados a temáticas tan diversas como asuntos internacionales (e.g. ingresos y egresos de personas al país), salud (e.g. casos de COVID-19 registrados en el país), educación (e.g. censos docentes) o medio ambiente (e.g. grado de degradación de tierras secas), entre muchos otros (dat20).

Los datos compartidos son aquellos que tienen licencias preferentes para casos de uso específicos (dat19). Además, de acuerdo al *Open Data Institute* o Instituto de datos abiertos (ODI) (the20), existen matices como los siguientes:

- Acceso con nombre: datos que solo se comparten con personas u organizaciones con nombre. Por ejemplo, los conjuntos de datos proporcionados bajo licencia para un propósito específico, como puede ser trabajar con estudiantes.
- Acceso basado en grupos: datos que están disponibles para grupos específicos que cumplen con ciertos criterios. Por ejemplo, los conjuntos de datos proporcionados bajo licencia como parte de servicios básicos para las partes interesadas.
- Acceso público: datos que están disponibles para cualquier persona bajo términos y condiciones que no son «abiertas». Por ejemplo, cualquier dato actualmente disponible al público en el *NBN Gateway*, una base de datos que contiene información de la biodiversidad del Reino Unido (nbn20).

Por otro lado, los datos cerrados son definidos por la ODI como datos a los que solo puede acceder su propietario o titular, que requieren una licencia personalizada por uso, negociada caso por caso (dat19).

En esta tesis se utilizaron datos abiertos de la Ciudad Autónoma de Buenos Aires generados en la plataforma «Gestión Colaborativa» (sol20). Esta plataforma está destinada a la atención eficiente de PQR realizados por los ciudadanos respecto a los servicios públicos brindados por la ciudad. También se hizo uso de datos cerrados de la misma índole, PQR reportados por los ciudadanos de San Isidro, Buenos Aires, mediante la plataforma Citymis (cit20). Estos datos fueron provistos por la empresa Mismática (mis20), previa sanitización, en un trabajo colaborativo junto con investigadores de la Universidad Nacional del Sur (UNS). El detalle de estos datos será abordado en el Capítulo 4.

2.2.2. Datos en redes sociales

Las redes sociales, como Twitter, Facebook y otras, proveen servicios que permiten establecer contacto entre personas por medio de una plataforma web. Mayormente, las redes sociales se utilizan para poder comunicarse con grupos de personas sin importar su ubicación geográfica y sin las limitaciones convencionales del correo electrónico o las llamadas telefónicas. Este espacio personal, y la anonimidad que otorgan, incentivaron a las personas a volcar sus opiniones sobre diversos temas. Además, la presencia de los representantes electos o de las instituciones gubernamentales en las mismas redes sociales permitieron acercar y permitir el contacto directo entre gobierno y ciudadano.

En general, los organismos gubernamentales optan por utilizar las redes sociales para crear una relación de comunicación unilateral con la ciudadanía, donde el gobierno comunica sobre las medidas tomadas, publicita o informa para concientizar a la población sobre determinada temática. Sin embargo, los ciudadanos siguen utilizando las redes sociales para hacerles llegar mensajes a sus representantes electos ya sea a favor o en contra de las medidas tomadas, y ya sea en un mismo anuncio del gobierno, por medio del uso de mecanismos propios de la red en cuestión (como, por ejemplo, utilizando hashtags) o en el espacio personal del ciudadano en una determinada red social.

Una red social muy popular a nivel global es Twitter (twi20), un servicio de microblogueo, identificado por permitir el envío de mensajes de texto plano de corta longitud (280 caracteres). La red permite *twittear* (publicar un mensaje), *retweetear* (compartir

un tweet publicado por otro usuario), responder a un tweet o marcarlo como favorito. Un usuario también puede observar eventos (nacionales e internacionales, como puede ser eventos deportivos o elecciones presidenciales) y tendencias, que son las temáticas (representadas por palabras o hashtags) más populares en un momento específico, determinado por un algoritmo que tiene en cuenta los gustos personales del usuario. Dada la naturaleza de esta red, Twitter es un canal que permite mostrar comentarios y opiniones de las personas sobre temas o cosas que suceden en diferentes partes del mundo.

Twitter también provee una interfaz de programación de aplicaciones (API)² mediante la cual se pueden obtener tweets utilizando palabras clave. El uso básico de esta API es gratis, pero limitado. Sin embargo, gracias a la facilidad de obtención de los tweets mediante la API y la naturaleza misma de Twitter, esta red social es muy popular en la comunidad científica interesada en analizar una base de datos de comentarios ciudadanos acerca de determinada temática.

Otra red social que ofrece facilidades para la obtención de datos es Reddit (red20). Reddit se define como una red de comunidades basada en los intereses de las personas. La misma ofrece una serie de *subreddits* (comunidades temáticas) donde los usuarios pueden compartir texto, imágenes, vídeos o enlaces, y votar a favor o en contra del contenido, que nivela la aparición de publicaciones más o menos destacadas. Esta red social es muy popular en países de habla anglosajona, y por lo tanto, gran cantidad de su contenido está en inglés, pero su popularidad va creciendo en el resto del mundo.

Reddit también ofrece una API³ mediante la cual se puede obtener información acerca de subreddits y recuperar publicaciones utilizando palabras clave, entre muchas otras opciones. Estas publicaciones pueden llegar a ser muy interesantes para analizar, sobre todo si se busca estudiar algún evento en países anglosajones, no solo porque es muy utilizada, sino porque suelen formarse cadenas de comentarios en respuesta a una determinada publicación, y, a diferencia de Twitter, las publicaciones no tienen límite de palabras.

En esta tesis se hizo uso de estas dos redes sociales, Twitter y Reddit, y el detalle de los datos utilizados será abordado en los Capítulos 5 y 6.

² API de Twitter - <https://developer.twitter.com/en/docs/twitter-api>

³ API de Reddit - <https://www.reddit.com/dev/api/>

2.2.3. Datos en periódicos digitales

El periodismo digital designa la modalidad de periodismo que tiene a Internet como entorno principal de desarrollo, incluyendo a las redes y dispositivos digitales en general. La principal virtud es la capacidad de inmediatez y actualidad que posee: todos los medios digitales pueden actualizarse a cualquier hora y desde cualquier lugar, siempre que el periodista o encargado tenga acceso a una computadora y a Internet. Además, los periódicos digitales pueden adaptarse mejor a las demandas y necesidades del público al tener la posibilidad de ofrecer contenido audiovisual como imágenes, vídeos y audios.

Los periódicos digitales ofrecen un espacio en cada noticia para una sección de comentarios, integrado al periódico en cuestión o mediante algún *plugin* de alguna red social, como por ejemplo Facebook. Esto ofrece una nueva forma para los ciudadanos de otorgar su visión, opinión, experiencia o conocimientos sobre un tema particular, ya sea en respuesta a una noticia de una determinada temática o al comentario de otro usuario en algún hilo de discusión.

En esta tesis se recuperaron noticias y comentarios asociados de periódicos digitales nacionales e internacionales, en particular de Clarín⁴, Página12⁵, LaNueva⁶ y CBC Canada⁷. El detalle de los datos utilizados será abordado en los Capítulos 5 y 6.

2.2.4. Otros

Si bien las redes sociales y los datos abiertos son las opciones más populares al momento de realizar un análisis de opiniones y participación ciudadana en la red, existen otras fuentes de información que pueden utilizarse para mejorar diferentes aspectos de instituciones gubernamentales relacionados a oír al ciudadano. Algunos requieren cierto trabajo, dado que es información o física y necesitan primero pasarse a datos digitales para ser posteriormente analizados. Ejemplos de fuentes de información de este estilo pueden ser encuestas realizadas por medio de llamadas telefónicas o censos poblacionales.

Diferentes encuestas y censos pueden estar ya digitalizados y accesibles desde portales de datos abiertos. También es posible que estén digitalizadas diferentes encuestas

⁴ Clarín - <https://www.clarin.com>

⁵ Página12 - <https://www.pagina12.com.ar>

⁶ LaNueva - <https://www.lanueva.com>

⁷ CBC Canada - <https://www.cbc.ca>

realizadas en programas de televisión o radio que cuentan con su propia página web. Estas encuestas realizadas a los oyentes suelen estar relacionados a temas de actualidad, sociedad o política, y requieren que los ciudadanos accedan a la página web para votar.

Así mismo, pueden analizarse diferentes discursos, anuncios o reportes; por ejemplo, discursos electorales o discusiones parlamentarias. Este tipo de información puede complementarse con las impresiones ciudadanas buscando adecuadamente en redes sociales o, de existir, en una plataforma pensado para ello.

Finalmente, otras fuentes de información dependen exclusivamente del desarrollo de una aplicación o plataforma para alguna iniciativa de participación ciudadana. Este tipo de plataformas permite la recopilación de información que puede ser posteriormente analizada y utilizada por funcionarios públicos.

2.3. Técnicas informáticas

El análisis de datos es definido como el proceso de inspección, limpieza, transformación y modelado de datos con el fin de descubrir información útil, informar conclusiones y respaldar la toma de decisiones. El proceso del análisis de datos va desde la identificación de la información requerida hasta la visualización final de los resultados, pasando por la recuperación, procesamiento, limpieza, análisis exploratorio, el modelado de los datos y la aplicación de algoritmos adecuados.

En esta sección se presentarán las técnicas para la recuperación (Sección 2.3.1), análisis (Sección 2.3.2) y visualización (Sección 2.3.3) de los datos. En particular, se ahondará en aquellas utilizadas en las herramientas desarrolladas en esta tesis.

2.3.1. Técnicas para la recuperación de datos

La ciencia de búsqueda y recuperación de información se basa en la búsqueda de datos en medios electrónicos con el objetivo de ser precisa y relevante. Para estas tareas generalmente se utilizan bases matemáticas, donde el modelo elegido determinará la forma en que se representan los documentos y las consultas realizadas por los usuarios, así como también el cálculo que se utiliza para definir la similitud entre ambos.

Algunos modelos de recuperación de información se basan en teoría de conjuntos, donde los documentos se representan como conjuntos de palabras y las similitudes se derivan de operaciones teóricas sobre conjuntos, como el modelo booleano estándar (*standard boolean model*) (BYRN11). Otros modelos utilizados son algebraicos, donde los documentos se representan como vectores o matrices y las similitudes son valores escalares, como el modelo de espacio vectorial (*vector space model*) (BYRN11). También existen modelos basados en probabilidad, donde el proceso de recuperación de documentos se trata como una inferencia probabilística, donde la similitud es la probabilidad de que un documento sea relevante para una dada consulta, como el modelo de independencia binario (*binary independence model*) (BYRN11). Todos estos modelos buscan maximizar la similitud entre las consultas realizadas por el usuario y los documentos que pueden obtenerse de la web, sin perder de vista la precisión (*precision*) y la exhaustividad (*recall*).

Para el desarrollo de esta tesis, se obtuvieron los datos en base a los siguientes métodos:

- Acceso a portales de datos abiertos: muchos datos están actualmente accesibles, por lo que solo requirió una búsqueda adecuada de los datos deseados dentro de los portales de datos abiertos de la ciudad de Buenos Aires.
- Comunicación con API: sitios tales como Twitter, Reddit y Facebook permiten a los desarrolladores solicitar información del sitio (como los posts o información pública de los perfiles) mediante la conexión a su API. Generalmente, esta conexión requiere la identificación del desarrollador y limita la cantidad de solicitudes por día (a excepción del pago de suscripción a sus servicios).
- Implementación de *crawlers* o *scrapers*: para los sitios que no proveen API, se puede desarrollar un *crawler* o *scraper* para extraer la información de manera automática.

Para las herramientas presentes en capítulos posteriores, se hizo uso de los datos del portal de datos abiertos de Buenos Aires, en particular aquellos datos relacionados al Sistema Único de Atención Ciudadana (SUACI)⁸. Además, se realizaron las conexiones con las API de Twitter y de Reddit para la obtención de tweets y posts respectivamente, y se hizo uso de librerías para la recuperación de las respuestas a los tweets, dado que la API de Twitter no provee ese servicio. Finalmente, se desarrollaron *scrapers* para recuperar la información de noticias y sus comentarios para una serie de periódicos digitales argentinos (Clarín, Página12 y LaNueva) y uno canadiense (CBC Canada)⁹.

⁸ SUACI - <https://data.buenosaires.gob.ar/dataset/sistema-unico-atencion-ciudadana>

⁹ newest - <https://github.com/rbhubert/newest>

Respecto a las tareas de preprocesamiento de estos datos (como la sanitización, limpieza y estructuración), estas se indicarán en los capítulos que describen las herramientas que hacen uso de estos datos.

2.3.2. Técnicas para el análisis de datos

Al realizar una revisión de la literatura, se identificaron el conjunto de técnicas para el análisis de datos más utilizado en las iniciativas de participación ciudadana presentes en artículos científicos. Estas son:

- **Analítica de datos:** incluye análisis cuantitativos y cualitativos de información que se puede utilizar en la toma de decisiones y en la identificación y resolución de problemas. Por ejemplo, la plataforma presentada en (VKIS16) utiliza la analítica web (usando número de visitas, visitantes, visitas a páginas y duración de la sesión, entre otros) para detectar eventos significativos en el sitio.
- **Crowdsourcing:** utiliza contribuciones de individuos externos para obtener ideas o servicios. Por ejemplo, los autores de (RY15) analizan dos experiencias de crowdsourcing relacionadas con el cambio climático. La plataforma utilizada permitió a las partes interesadas proponer nuevas medidas sobre el tema en consideración.
- **Gamificación:** se refiere al uso de elementos de juego (como puntos y niveles) para mejorar la experiencia del usuario y motivar la participación en una aplicación que no es un juego (DSN⁺11). Muchas plataformas utilizan esta técnica para involucrar a los ciudadanos en la realización de propuestas al municipio (como informar un problema o proponer una iniciativa) recompensándolos con puntos (Ber12; BFR16).
- **Minería de datos:** esta categoría incluye todas las técnicas que se utilizan para extraer patrones y conocimiento de grandes conjuntos de datos, utilizando métodos que involucran aprendizaje automatizado, estadística y sistemas de base de datos (HA14). Las dos técnicas más representativas son el agrupamiento y la clasificación: la primera permite descubrir grupos y estructuras similares en los datos, y la segunda permite identificar la categoría de un nuevo elemento dado un conjunto inicial de clases y ejemplos de entrenamiento.

Dentro de esta categoría destacamos la minería de opinión como una subcategoría.

- Minería de opinión: busca recopilar y categorizar opiniones sobre un tema en particular utilizando procesamiento de lenguaje natural, análisis de texto y análisis de sentimiento, entre otras técnicas. Por ejemplo, en (SCB⁺10) se propone un framework para detectar y analizar automáticamente la opinión pública sobre las decisiones gubernamentales.
- Modelado de simulación: esta técnica implica crear y analizar un prototipo digital de un modelo físico para predecir su desempeño en el mundo real. Por ejemplo, una plataforma que utiliza modelos de simulación para estimar los resultados de diversas propuestas ciudadanas sobre políticas públicas (SCLD14).
- Realidad aumentada: esta tecnología permite la modificación de la realidad, ya sea generando o extrayendo información sensorial del mundo real, como sonido o vídeo. Por ejemplo, en (RBS⁺14) se utiliza realidad aumentada para mantener informados a los ciudadanos sobre la planificación urbana.
- Sistema de información geográfica (GIS): son tecnologías utilizadas para asociar varios tipos de elementos con sus coordenadas geográficas. Esta información puede ser usada para realizar análisis espaciales o visualizar componentes en un mapa. Por ejemplo, en (KH14a) se presenta una iniciativa que utiliza GIS para visualizar reportes ciudadanos en un mapa de la ciudad.
- Tecnologías semánticas: estas tecnologías utilizan ontologías¹⁰ para proporcionar una expresión que explore el significado de los recursos (como contenido, personas o sistemas, entre otros) en un formato legible por una computadora (TDMW09). Por ejemplo, la plataforma presentada en (MP14) extrae información semántica de los comentarios de los ciudadanos para realizar análisis de opinión.

A continuación, se detallarán las técnicas y algoritmos que se utilizaron en las herramientas desarrolladas en esta tesis: análisis de agrupamiento (Sección 2.3.2.1), análisis de sentimiento (Sección 2.3.2.2) y aprendizaje activo (Sección 2.3.2.3).

¹⁰En (Gru93) se define ontología como una especificación explícita de una conceptualización. Una ontología comprende una representación y definición de las categorías, propiedades y relaciones entre conceptos, datos y entidades de un dominio.

2.3.2.1. Análisis de agrupamiento

Uno de las técnicas más representativas de la minería de datos es el *clustering* (o agrupamiento) que consiste en la agrupación un conjunto de objetos de tal forma que aquellos en un mismo clúster (o grupo) son más similares entre sí que con aquellos que pertenezcan a otro clúster (Ber06). Existen diferentes algoritmos para la tarea de agrupamiento que difieren significativamente en su comprensión de lo que constituye un clúster y cómo encontrarlos de manera eficiente. Algunas nociones de clústeres incluyen grupos con pequeñas distancias entre los objetos, áreas densas del espacio de datos, e intervalos o distribuciones estadísticas particulares.

Cada algoritmo de agrupamiento puede ser clasificado dentro de dos grandes técnicas para agrupamiento de objetos:

- Agrupamiento jerárquico: busca construir una jerarquía de clústeres. Hay dos estrategias para este agrupamiento, aglomerativa (*bottom-up*), donde cada objeto es un clúster y los pares de clústeres son mezclados a medida que se sube en la jerarquía, y divisivas (*top-down*), donde todos los objetos conforman un único clúster y se realizan divisiones a medida que uno baja en la jerarquía.

Estos algoritmos son capaces de definir por sí solos el número de clústeres. En general, las mezclas o divisiones son determinadas con un algoritmo *greedy*.

- Agrupamiento no jerárquico: aquí la relación entre los clústeres no está determinada, no sigue una estructura de árbol como en el agrupamiento jerárquico. En esta técnica se agrupan los datos con el fin de maximizar o minimizar cierto criterio de evaluación.

Dentro de la técnica de agrupamiento jerárquico se encuentra el **algoritmo *Hierarchical greedy clustering*** o «agrupamiento jerárquico codicioso». Este algoritmo es utilizado para lograr la visualización de agrupaciones de la manera más eficiente posible, dado que su rendimiento es lo suficientemente bueno como para ejecutarse dinámicamente incluso con un gran número de puntos.

El algoritmo de agrupamiento codicioso funciona de la siguiente manera:

1. Se selecciona un punto del conjunto de datos.
2. Se encuentran todos los puntos cercanos al elegido en un determinado radio.
3. Se crea un nuevo clúster con los puntos encontrados.
4. Se elige un nuevo punto que no forma parte de ningún clúster y se repiten los pasos hasta haber visitado todos los puntos en el conjunto de datos.

Este algoritmo puede utilizarse para encontrar agrupaciones de puntos a un determinado nivel de zoom en un mapa de alguna ciudad, que es uno de los usos que se le da en esta tesis. Sin embargo, sería muy costoso hacerlo para todos los niveles de zoom permitidos. Por ejemplo, si los niveles de zoom van de 0 a 15, se debería procesar todo el conjunto de datos 16 veces, uno por cada nivel; además, realizar agrupamiento en los niveles inferiores de zoom tardaría demasiado dado que cada agrupación tendría que alojar a un número exponencialmente creciente de puntos.

Este problema puede evitarse reutilizando evaluaciones previas para calcular las nuevas agrupaciones. Luego de realizar agrupamiento sobre un dado nivel de zoom (por ejemplo, el zoom 15), se pueden agrupar los clústeres resultantes para determinar los clústeres del zoom inmediato inferior (siguiendo el ejemplo, del zoom 14) y así sucesivamente. Dado que el número de puntos a analizar decrece exponencialmente en cada paso, como puede verse en la Figura 2.1, se puede realizar agrupamiento en todos los niveles mucho más rápido. Es a este enfoque al que se le llama *hierarchical greedy clustering* (Map18).

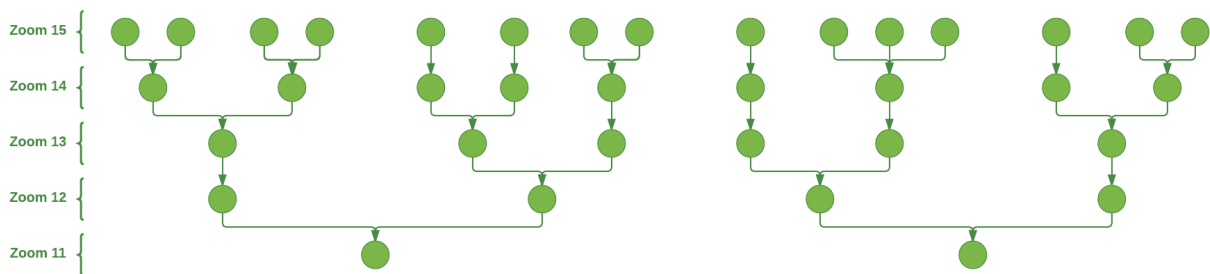


Figura 2.1: Ejemplo de *hierarchical greedy clustering*.

Por otro lado, dentro de la técnica de agrupamiento no jerárquico, se encuentra el **DBSCAN** o «agrupamiento espacial basado en densidad de aplicaciones con ruido» (EKS⁺96). En este algoritmo los clústeres se definen como áreas de mayor densidad respecto al resto del conjunto de datos. Los objetos en estas áreas de escasas densidad (que separan los clústeres) son considerados puntos de ruido y puntos de frontera.

En el caso del algoritmo DBSCAN, los puntos pueden clasificarse como *core points* (puntos núcleos), *(density-) reachable points* (puntos -densamente- alcanzables) y *outliers* (puntos ruidos) de la siguiente manera:

- Un punto p es un punto núcleo si al menos $minPts$ puntos están a una distancia ϵ de él. Estos son los puntos *directamente alcanzables* desde p . Ningún punto puede ser *directamente alcanzable* desde un punto que no es núcleo.
- Un punto q es *alcanzable* si existe un camino de puntos entre p y q donde cada punto en ese camino es un punto núcleo (con la posible excepción de q).
- Todos los puntos no alcanzables desde cualquier otro punto se consideran *outliers*.

Si un punto p es un punto núcleo, conforma un clúster con todos aquellos puntos que son alcanzables desde él. De este modo, cada clúster contiene al menos un *core point*; los puntos que no son núcleo pueden ser parte del grupo, pero son considerados parte de la «periferia» dado que a partir de ellos no pueden alcanzarse nuevos puntos.

La relación de «ser alcanzable» no es simétrica dado que ningún punto puede ser alcanzable desde un punto que no es núcleo (es decir, un punto no-núcleo puede ser alcanzable, pero no se puede alcanzar desde él). Por lo tanto, se necesita otra noción de conectividad para definir formalmente la extensión de un clúster encontrado por el algoritmo: dos puntos p y q están *densamente conectados* si existe un punto o tal que tanto p como q son (densamente) alcanzables desde o . Esta relación es simétrica.

Un clúster encontrado con DBSCAN satisface dos propiedades:

1. Todos los puntos de un mismo clúster están densamente conectados.
2. Si un punto es densamente alcanzable desde cualquier otro punto en ese clúster, entonces también forma parte del clúster.

El algoritmo DBSCAN requiere de dos parámetros para su ejecución:

- ϵ (épsilon), que funciona como el valor de distancia requerido para determinar la cercanía entre puntos.
- $minPts$, la mínima cantidad de puntos que son requeridos para que una región se considere densa.

El algoritmo funciona de la siguiente manera:

1. Se comienza con un punto p arbitrario que no haya sido visitado.
2. Se computa el vecindario de puntos de p en un radio ϵ de distancia.
3. Si el vecindario contiene los puntos suficientes ($minPts$), se crea un nuevo clúster. De forma contraria, el punto se considera *outlier*. Es importante destacar que este punto puede luego encontrarse en el vecindario de otro punto y ser considerado parte de otro clúster.

4. Si un punto se incluye en la parte densa de un clúster, su ϵ -vecindario también es parte de ese clúster, siguiendo las definiciones antes mencionadas. De este modo, todos los puntos encontrados en la ϵ -vecindad son agregados, así como también sus propios ϵ -vecindarios cuando ellos también son densos. Este proceso continúa hasta que el clúster densamente conectado es encontrado en su totalidad. Luego, se vuelve al paso 1 hasta visitar todos los puntos.

En general, las implementaciones de este algoritmo requerirán como parámetros de entrada, además del conjunto de datos inicial, *epsilon* y *minPts*. Opcionalmente, también puede elegirse la función de distancia utilizada por el algoritmo y el modelo geográfico.

2.3.2.2. Análisis de sentimientos

El análisis de sentimientos, también conocido como minería de opinión, se refiere al uso de procesamiento de lenguaje natural y aprendizaje automatizado para interpretar y clasificar estados afectivos, polaridad o emociones en datos subjetivos, como pueden ser textos en redes sociales. Los modelos de análisis de sentimiento pueden enfocarse en polaridad (positivo, negativo, neutro), así como también en emociones (felicidad, enojo, tristeza, etc.) o en intenciones (interesado o no interesado).

Los principales algoritmos para el análisis de sentimientos incluyen:

- Algoritmos basados en reglas: se realiza análisis de sentimientos basado en un conjunto de reglas creadas manualmente.
- Algoritmos automatizados: se basa en el uso de técnicas de aprendizaje automatizado para aprender de los datos.
- Algoritmos híbridos: estos sistemas combinan los dos enfoques antes mencionados.

Independientemente del enfoque elegido, se puede optar por aplicar una serie de técnicas de preprocesamiento de los datos, como pueden ser:

- *Stemming*: consiste en reducir una palabra a su raíz (en inglés, a un *stem*).
- *Tokenization*: se busca dividir cadenas de texto largas en piezas más pequeñas o *tokens*, donde cada uno de estos *tokens* tiene un significado asignado. Por ejemplo, se puede dividir una oración en palabras, cada palabra será un *token*.
- *Part-of-speech tagging*: consiste en etiquetar las palabras en un texto con su categoría gramatical, como puede ser adjetivo, sustantivo, verbo, etc.

- *Parsing*: es el proceso de analizar una cadena de símbolos en base a las reglas de una gramática formal.

El **enfoque basado en reglas** utiliza un conjunto de reglas creadas manualmente que ayudan a identificar la subjetividad, polaridad o el tema de una opinión. Generalmente hacen uso de un lexicon, que consiste en una lista de palabras o expresiones y la polaridad o emociones asociadas. Un sistema basado en reglas básico funciona de la siguiente manera:

1. Se definen dos listas de palabras polarizadas (por ejemplo, las palabras clasificadas como negativas pueden ser «malo», «peor», «feo», mientras que las palabras positivas pueden ser «bueno», «mejor», «bello»).
2. Utilizando estas listas, se cuentan el número de palabras positivas y negativas que aparecen en un determinado texto.
3. Si el número de palabras positivas es mayor al número de palabras negativas, el sistema determina que ese texto tiene un sentimiento positivo asociado, y viceversa. Si el número de apariciones de palabras positivas y de negativas es el mismo, el sistema determina un sentimiento neutral.

Esta clase de sistemas son ingenuos considerando que no tienen en cuenta cómo las palabras se combinan en el texto. Se pueden utilizar técnicas de procesamiento más avanzadas y agregar nuevas reglas para dar respaldo a nuevas expresiones y vocabulario. Sin embargo, agregar nuevas reglas puede afectar a los resultados anteriores, y el sistema puede volverse muy complejo rápidamente.

El **enfoque automatizado** se basa en el uso de algoritmos de aprendizaje automatizado. Se modela normalmente como un problema de clasificación, donde dado un texto se espera recibir una categoría (por ejemplo, positiva, negativa o neutral).

Generalmente un problema de este tipo tiene dos procesos principales:

- Proceso de entrenamiento: donde el modelo aprende a asociar determinada entrada (el texto) con el valor de salida correspondiente (la categoría), basado en un conjunto de ejemplos utilizado para el entrenamiento.

Este proceso tendrá un subproceso que se encargará de extraer características del texto para crear un vector de características. El algoritmo de aprendizaje automatizado recibe pares de vector de características y categorías para generar el modelo.

- Proceso de predicción: se utiliza el mismo extractor de características para transformar entradas nuevas en vectores de características. Con estos vectores, el algoritmo generará una predicción de la categoría para ese texto.

Como hemos visto, el primer paso en un clasificador es transformar el texto en un vector de características. La representación más básica es *bag-of-words* o *bag-of-ngrams* con su frecuencia. Otras técnicas de extracción de características más sofisticadas están basadas en *word embeddings*, representaciones que hacen posible que palabras con significados similares tengan representaciones similares, que eventualmente pueden mejorar el rendimiento de los clasificadores.

Los algoritmos de clasificación utilizados involucran modelos estadísticos como el *Naïve Bayes*, *Logistic Regression*, *Support Vector Machines* y *Neural Networks*. En particular, un método utilizado en esta tesis es la «Regresión logística multinomial» (*Multinomial Logistic Regression*), que generaliza el método de regresión logística para problemas multiclase, es decir, con más de dos posibles resultados discretos. El modelo se utiliza para predecir las probabilidades de los diferentes resultados posibles de una distribución categórica como variable dependiente dado un conjunto de variables independientes.

2.3.2.3. Aprendizaje activo

El *active learning* o aprendizaje activo es un caso especial de aprendizaje automatizado donde el algoritmo puede consultar al usuario (o a otra fuente de información) para etiquetar nuevos datos con los resultados deseados (Set09).

Estos algoritmos son muy útiles en situaciones donde los datos no-etiquetados son demasiados y el trabajo de etiquetarlos manualmente es muy costoso. En estos casos, se consulta al experto para que etiquete unos pocos ejemplos muy informativos para el algoritmo, de modo que aprenda más rápido con menos ejemplos etiquetados. Este proceso es iterativo hasta que se alcanzan los resultados deseados.

Este proceso iterativo se puede plantear de la siguiente manera. Sea T el conjunto de todos los puntos de datos que se están considerando. En cada iteración, T va a estar dividido en tres categorías:

- T conocido: los puntos de datos cuya categoría es conocida.
- T desconocido: los puntos de datos cuya categoría se desconoce.

- T elegidos: un subconjunto de T desconocido que son elegidos para ser etiquetados por el experto.

Los algoritmos para determinar qué puntos de datos deberán ser etiquetados por el experto pueden organizarse en diferentes categorías dependiendo de su propósito. Por ejemplo aquellos que más cambiarán el modelo actual, los que más reducirán el error de generalización del modelo o la varianza de la salida, aquellos para los que el modelo actual tiene menos certeza de cuál debería ser la salida correcta, entre otros.

2.3.3. Técnicas para la visualización de datos

La visualización de datos es la representación gráfica de los datos de manera tal que las imágenes producidas comuniquen las relaciones entre los datos representados. Una visualización eficaz ayuda a los usuarios a analizar y razonar sobre los datos y las pruebas que se realicen, haciendo que datos complejos sean más accesibles, comprensibles y utilizables. Por lo tanto, una correcta visualización de los datos es esencial para la interpretación, evaluación y posterior toma de decisiones sobre los datos evaluados.

Existen diversas técnicas de visualizaciones que dependen de los datos que quieran representarse. A continuación se describirán técnicas de visualizaciones dependiendo del tipo de dato a visualizar: geográficos o espaciales, cuantitativos, categóricos u otros.

2.3.3.1. Datos geográficos o espaciales

Para datos que contienen ubicación geográfica (como latitud y longitud, o calle y altura) se pueden utilizar visualizaciones de mapas de la ciudad a los que se les agrega marcadores, como líneas, puntos o polígonos, entre otros. Por ejemplo, al tratar con reclamos ciudadanos sobre problemas urbanos, los reportes pueden representarse como puntos sobre el mapa. Si se realizan otros análisis sobre estos datos, las visualizaciones pueden volverse tan complejas como sea requerido, como por ejemplo crear un mapa de calor.

Un **mapa de calor** es una técnica que muestra la magnitud de un determinado fenómeno con colores. La ausencia o presencia de colores y su nivel de intensidad está asociada a la presencia o agrupación de los fenómenos en ese espacio, como puede ser, por ejemplo, muchos pedidos para el arreglo de una calle. La variación de color puede ser de tono o de intensidad, y es esta variación lo que proporciona información relacionada a

cómo se agrupan los datos o cómo varían en el espacio. Generalmente se emplean colores fríos para indicar baja actividad y colores cálidos para indicar alta actividad.

En un mapa también se pueden dibujar puntos y líneas, por ejemplo, para representar un camino; o polígonos que se pueden utilizar, por ejemplo, para marcar los límites geográficos de un conjunto de problemáticas reportadas por ciudadanos en una ciudad. También se pueden distorsionar aspectos de un mapa para acentuar datos clave (cartograma), o utilizar diferentes tonos de una misma gama cromática para representar distintos valores de una variable estadística, como la densidad poblacional (mapa coroplético). Hay muchas variables que se pueden utilizar sobre un mapa, normalmente jugando con marcadores (como líneas, puntos o polígonos) y colores.

2.3.3.2. Datos cuantitativos

La mayoría de las visualizaciones disponibles se utilizan para comunicar información cuantitativa. Tal vez el más básico, y común para diferentes campos de aplicación, es el *line chart* o **gráfico de líneas**, un tipo de gráfico que muestra información como una serie de puntos de datos (llamados marcadores) conectados por líneas rectas. Este tipo de gráfico generalmente se utiliza para visualizar datos que varían con el tiempo (es decir, un *time series* o serie de tiempo).

Un diagrama similar es el *scatter plot* o **diagrama de dispersión**, un tipo de gráfico que utiliza las coordenadas cartesianas para representar valores para dos variables en el conjunto de datos. Adicionalmente, si los puntos están codificados (ya sea por color, tamaño o forma), se puede mostrar una variable adicional. En este gráfico, los datos se muestran como una colección de puntos, cada uno con el valor de las variables que determinarán su posición en el eje vertical y horizontal. Una variación del diagrama de dispersión es el *bubble chart* o **gráfico de burbujas**, donde en lugar de puntos se tienen burbujas. Este diagrama se utiliza para mostrar tres dimensiones de los datos, las primeras dos mediante su posición *xy* y la tercera mediante su tamaño. Otro diagrama similar es el *event drop diagram*, que permite visualizar un aspecto medible de los datos sobre una línea de tiempo. En el diagrama se presentan líneas horizontales que representan tiempo, y sobre ellas pueden dibujarse puntos o burbujas, cuyo tamaño representa el valor de un dado evento en un punto de tiempo en particular.

Otro diagrama muy utilizado es el **histograma**, que es una representación aproximada de la distribución de datos numéricos. En este diagrama primero se divide todo el rango

de valores en una serie de intervalos, a los que llamaremos *bins* o recipientes, y luego se cuentan la cantidad de puntos que pertenecen a cada intervalo. Esto se representará luego usando rectángulos: cada rectángulo es un compartimiento y su altura es la cantidad de puntos en él. No debe confundirse este diagrama con el gráfico de barras, ya que el histograma se utiliza para variables continuas mientras que el gráfico de barras se utiliza para variables categóricas.

Finalmente, para datos con múltiples variables se puede utilizar un *radar chart* o **gráfico de radar**, que consta de una secuencia de radios equiangulares, donde cada radio representa una variable. El valor representado en cada radio será proporcional a la magnitud de la variable para un punto de datos en relación a la máxima magnitud de la variable para todos los puntos de datos; además, una línea conecta los valores de los datos de cada radio. Un gráfico equivalente al anterior son las **coordenadas paralelas**. En este gráfico se pueden mostrar un conjunto de datos en un espacio n-dimensional utilizando n líneas paralelas, típicamente verticales e igualmente espaciadas. Un punto en el espacio n-dimensional se representa como una polilínea con vértices en los ejes paralelos.

2.3.3.3. Datos categóricos

Cuando se tienen datos categóricos, se puede utilizar un *bar chart* o **gráfico de barras**. Este gráfico representa estos datos con categorías discretas mediante barras rectangulares, cuya altura será proporcional al valor que representa. Este gráfico puede mostrar las barras vertical u horizontalmente, y otras variaciones permiten agrupar barras en grupos o apilarlas, mostrando los valores de varias variables. Otra visualización básica muy utilizada es el *pie chart* o **gráfico de torta**, un gráfico estadístico circular, que se divide en porciones. La longitud del arco de cada rebanada (es decir, su ángulo y área) es proporcional a la cantidad que representa.

Si se tienen datos jerárquicos, una buena visualización es *sunburst diagram* o **gráfico de proyección solar**, que representa estos datos mediante círculos concéntricos. El círculo en el medio representa el nodo raíz, con la jerarquía moviéndose hacia afuera del centro. El nodo raíz se divide en secciones que van a ir conformando ramas. Cada una de estas secciones puede dividirse en subsecciones tanto como se quiera, llegando finalmente a las subsecciones «hojas». El área de cada sección será correspondiente al valor de esa sección (o tipo de dato), y los valores de las subsecciones sumarán el valor total para la

sección a la que pertenecen. Esta visualización permite observar cómo los círculos o anillos más exteriores se relacionan con los interiores.

Otra forma de representar datos categóricos es mediante el **diagrama de Venn** o el **diagrama de Euler**, que permite representar conjuntos y sus relaciones. Son útiles para explicar jerarquías complejas y definiciones superpuestas. La diferencia entre diagramas de Venn y de Euler es que el primero representa todas las posibles relaciones entre los conjuntos, mientras que los diagramas de Euler representan las intersecciones no vacías.

Un diagrama diferente para mostrar relación entre conjuntos es el *extended Hasse diagram* o **diagrama de Hasse extendido** (propuesto en (HEMJ18)), un diagrama que permite visualizar relaciones entre conjuntos, donde el tamaño de los nodos es proporcional al tamaño de los conjuntos. Esta técnica es adoptada para analizar combinaciones de diferentes tipos de contribuciones ciudadanas (comentarios, retweets y favoritos).

Finalmente, el **diagrama de Sankey** es un tipo de diagrama de flujo donde el ancho de las flechas es proporcional al volumen de los flujos. Si bien parece un tipo de visualización específica para algunos campos de la ciencia, como por ejemplo física, se puede adaptar para representar otro tipo de relaciones.

2.3.3.4. Otros tipos de datos

Para representar un tipo de dato de texto, como documentos, se puede utilizar una *word cloud* o **nube de palabras**, que es una representación visual de las palabras que conforman un texto, donde el tamaño de las palabras está asociado a su frecuencia en el texto. Otra visualización puede ser un *word tree* o **árbol de palabras**, que muestra los datos de texto de forma jerárquica, como un árbol de palabras individuales conectados por líneas. El tamaño de la palabra representa su peso, ya sea la frecuencia o la cantidad de «hijos» en el árbol. Este tipo de visualización se puede utilizar para mostrar una jerarquía de términos o los contextos típicos en los que aparece una palabra.

Otro tipo de datos puede requerir visualizaciones específicas. Esto generalmente está asociado a sistemas o aplicaciones complejas específicas a determinadas ciencias, como puede ser una visualización para observar los diferentes estratos de piedras o tierra en aplicaciones relacionadas a geología, o visualizar la composición de moléculas o compuestos químicos en una aplicación relacionada a química o biología. Una de las visualizaciones identificadas en la revisión de literatura es aquella para la visualización de argumentos

para simplificar su interpretación y análisis. Por ejemplo, en (BM12) se presentan mapas de debate navegables con opiniones de los interesados sobre las políticas públicas.

2.4. Resumen

En este capítulo se presentaron los conceptos transversales al contenido de la tesis. Se comenzó presentando el concepto de participación electrónica y las iniciativas de participación ciudadana, indicando la existencia de diversos niveles y áreas de e-Participación, y de origen de las iniciativas, pudiendo ser de cualquier sector, con mayor o menor participación del gobierno y los ciudadanos. A continuación, se presentaron las diversas fuentes de información que pueden utilizarse al momento de crear nuevas iniciativas, incluyendo datos abiertos y cerrados, y aquellos presentes en las redes sociales y en periódicos digitales. Finalmente, se presentó un conjunto de diferentes técnicas informáticas que pueden utilizarse tanto para recuperar información como para analizarla e incluso visualizarla, siempre teniendo en cuenta el tipo de dato que se está utilizando.

Capítulo 3

Trabajos Relacionados

En este capítulo se presentará el análisis sobre trabajos relacionados, particularmente respecto a aquellos que permiten identificar los aspectos relevantes de la participación electrónica (Sección 3.1) y las principales técnicas informáticas que se utilizan en iniciativas de participación ciudadana (Sección 3.2). A continuación, se realiza una evaluación del estado del arte con los trabajos identificados previamente (Sección 3.5). Finalmente, en la Sección 3.6 se ofrece un resumen del capítulo.

3.1. Participación electrónica

En esta sección discutiremos los trabajos relevantes de la literatura con el objetivo de mostrar los diferentes aspectos de e-Participación. En particular, presentaremos trabajos que identifican diferentes dimensiones de e-Participación, las técnicas utilizadas, y los principales problemas y desafíos al desarrollar iniciativas.

Varios trabajos estudian la e-Participación y definen diferentes dimensiones para analizar y clasificar iniciativas. En (TMC⁺07) se definen diferentes parámetros para entender qué es e-Participación: nivel, área, herramientas y barreras; estos parámetros se utilizan luego para clasificar las iniciativas de participación ciudadana en (TKT08). En (Mac04; Wim07) se identifican otras dimensiones para caracterizar una iniciativa de e-Participación, específicamente la etapa en el proceso de formulación de políticas, actores, reglas de compromiso, duración y sostenibilidad, accesibilidad, recursos y promoción, evaluación y resultados, y factores críticos de éxito. Los autores de (Med11) presentan un

análisis longitudinal del desarrollo de iniciativas de e-Participación basada en un framework que identificó, analizó y clasificó iniciativas teniendo en cuenta los actores, actividades, factores contextuales, efectos y evaluación. El estudio en (WDB18) analiza la literatura para unificar la definición de e-Participación, y propone un marco integrado que combine factores estratégicos y organizacionales, utilizando las siguientes categorías para clasificar una iniciativa: actores, objetivos, niveles, estrategias y herramientas. En (BML16) se examina el estado actual del arte sobre e-Participación, analizando dominios, desafíos, herramientas y sus objetivos. El trabajo en (SC19) tiene como objetivo analizar la relación entre las plataformas de e-Participación, las posibilidades de participación de esas iniciativas y el impacto en el proceso de toma de decisiones. Los autores clasificaron las plataformas en diferentes niveles de e-Participación en función de las funcionalidades y el nivel de interacción entre ciudadanos y gobierno.

Otros trabajos analizan iniciativas de participación electrónica para identificar los principales problemas y desafíos que se enfrentan al implementarlas. En (MMD17) se define una tipología de temas relacionados a la implementación de procesos de participación ciudadana, incluyendo seis categorías: ética, eficiencia y rentabilidad, política, calidad, ciudadanos y tecnología. El trabajo (HBML17) se centra en iniciativas de e-Participación mediante el examen de las motivaciones, acciones y percepciones públicas, identificando barreras a la e-Participación y evaluando las diferentes funciones y aplicaciones de las TIC para la participación ciudadana en la sustentabilidad ambiental. En (OG19) se presenta una revisión de la literatura para responder cuáles son las razones de la baja participación ciudadana (identificando barreras técnicas y no técnicas), cuál es el papel del gobierno en la e-Participación y cuáles son los enfoques actuales para promover e-Participación. La revisión presentada en (SSS19) se centra en las administraciones públicas y tiene como objetivo estructurar y sistematizar la literatura sobre las fases de difusión de la e-Participación y los niveles de análisis, identificando una serie de barreras o desafíos como la percepción pública, los costos y el diseño de e-Participación, entre otros. En (ARS⁺17) se presenta un análisis del papel de las redes sociales en los servicios de e-Gobierno para los ciudadanos, identificando una serie de cuestiones clave, como la motivación pública, la transparencia, la comunicación, la confianza y la seguridad, entre otros.

Algunos trabajos relacionados se centran en el uso de diferentes técnicas en iniciativas de participación electrónica. El trabajo presentado en (BC12) estudia el uso de tecnologías para la e-Participación usando una taxonomía que clasifica diferentes tipos de visión:

Tabla 3.1: Principales aspectos de las iniciativas de e-Participación identificadas en la revisión de la literatura.

| Tema | Categorías | Referencias |
|--|--|---|
| Dimensiones | Niveles | (TMC ⁺ 07; Mac04; Wim07; SPCG19; SC19; BML16; WDB18) |
| | Áreas | (TMC ⁺ 07; Wim07; SPCG19; Med11; EMT11; ZLP09) |
| | Herramientas o tecnologías | (TMC ⁺ 07; Mac04; Wim07; SPCG19; BML16; EMT11; ZLP09; WDB18) |
| | Barreras, desafíos o problemas | (TMC ⁺ 07; OG19; BML16; SSS19) |
| | Etapas en el proceso de formulación de políticas | (Mac04; Wim07) |
| | Actores | (Mac04; Wim07; SPCG19; Med11; WDB18) |
| | Reglas del compromiso | (Mac04) |
| | Duración y sostenibilidad | (Mac04) |
| | Accesibilidad | (Mac04) |
| | Recursos y promoción | (Mac04) |
| | Evaluación y resultados | (Mac04; Med11) |
| | Factores críticos de éxito | (Mac04) |
| | Métodos o estrategias | (SPCG19; WDB18) |
| | Procesos | (SPCG19) |
| | Roles | (SPCG19; OG19) |
| Factores contextuales | (Med11) | |
| Efectos de la e-Participación | (Med11) | |
| Dominio u objetivos | (BML16; PK08; WDB18) | |
| Barreras, desafíos o problemas | Políticos-estratégicos | (TMC ⁺ 07; MMD17) |
| | Organizacional y legal | (TMC ⁺ 07) |
| | Rol de la e-Participación | (TMC ⁺ 07; SSS19) |
| | Social | (TMC ⁺ 07; MMD17; SSS19; Kin04) |
| | Relacionado a la tecnología | (TMC ⁺ 07; MMD17; HBML17; OG19) |
| | Producción | (TMC ⁺ 07) |
| | Éticos | (MMD17) |
| | Eficiencia y rentabilidad | (MMD17; SSS19) |
| | Calidad | (MMD17) |
| | Motivaciones públicas | (HBML17; OG19; SSS19; ARS ⁺ 17) |
| | Percepciones públicas | (HBML17; OG19; SSS19) |
| | Privacidad/Seguridad | (OG19; ARS ⁺ 17) |
| | Transparencia | (OG19; ARS ⁺ 17) |
| Comunicación, interacción o colaboración | (ARS ⁺ 17) | |
| Confianza | (ARS ⁺ 17) | |
| Herramientas o tecnologías | Relacionadas con la web | (TMC ⁺ 07; Mac04; Wim07; SPCG19; BML16; BC12; EMT11; PK08; ZLP09; BTRF12; WDB18) |
| | Redes sociales | (TMC ⁺ 07; SPCG19; BML16; BC12; ZLP09; BTRF12; ARS ⁺ 17) |
| | Gamificación | (TMC ⁺ 07; Wim07; EMT11; Han07; ZLP09) |
| | e-Paneles | (TMC ⁺ 07; Wim07; EMT11; ZLP09) |
| | e-Peticiones | (TMC ⁺ 07; Mac04; Wim07; SPCG19; BML16; EMT11) |
| | Sistemas de información geográfica | (TMC ⁺ 07; Wim07; SPCG19; BC12; EMT11; Han07) |
| | e-Encuesta / e-Votación | (TMC ⁺ 07; Wim07; SPCG19; BML16; EMT11) |
| | Visualizaciones | (Mac04; SPCG19; BC12; EMT11; PK08; WDB18) |
| | Sistemas argumentativos | (Wim07) |
| | Ontologías | (Wim07) |
| | Plataformas para e-Consultas | (SPCG19; EMT11; WDB18) |
| | Software colaborativo | (SPCG19; WDB18; Han07) |
| | Procesamiento de lenguaje natural | (BC12; EMT11; WDB18) |
| | Crowdsourcing | (BC12; ARS ⁺ 17) |
| | Analítica de datos | (BC12) |
| Minería de datos | (BC12) | |
| Realidad aumentada | (Han07) | |

visiones integradoras y visiones de gobierno. En (EMT11) el objetivo fue explorar el estado del arte para analizar el uso de herramientas y tecnologías en diferentes áreas de e-Participación, identificando herramientas como chat o foros en línea, e-Polling, sistemas

de información geográfica, entre otros. Los autores de (BTRF12) analizan el uso de la Web 2.0 y las herramientas de redes sociales en los gobiernos locales de la Unión Europea. El análisis de tecnologías tales como *widjets*, *blogs* y *wikis* se utiliza para determinar su utilidad y si se están usando para incrementar la transparencia y la e-Participación.

Otros trabajos proponen clasificaciones de tecnologías basadas en un conjunto de métricas. El trabajo presentado en (PK08) realiza un estudio acerca del uso de TIC en iniciativas de e-Participación teniendo en cuenta los objetivos de e-Participación, como pueden ser el sondeo de insumos, la educación y el intercambio de información. Los autores de (Han07) presentan una revisión de tecnologías que pueden ser usadas en la planificación urbana, como los sistemas de información geográfica, las tecnologías de juego, el software colaborativo y la realidad aumentada. En (ZLP09) se presenta una comparativa entre las herramientas utilizadas en e-Participación en aspectos como seguridad, privacidad, accesibilidad y punto de vista del usuario. Además provee un análisis de Fortalezas, Oportunidad, Debilidades y Amenazas (FODA) para cada herramienta.

Un resumen de todas las dimensiones, barreras y desafíos, y tecnologías o herramientas identificadas en los trabajos antes mencionados puede verse en la Tabla 3.1.

3.2. Técnicas informáticas en iniciativas de participación electrónica

A fin de responder la pregunta de investigación definida previamente en la Sección 1.2: **¿Cuáles son las técnicas de minería de datos y de visualización más convenientes para el análisis de opiniones, quejas y reclamos vertidos en diversas plataformas digitales en el contexto de participación electrónica?**, se ha realizado una revisión de literatura que busca identificar iniciativas de e-Participación para definir una clasificación que tiene en cuenta el nivel, área, origen y técnicas utilizadas.

En este trabajo se comenzó con una búsqueda inicial de artículos de investigación utilizando términos como «herramientas» y «e-Participación», y se seleccionaron aquellos que presentan una iniciativa, plataforma o aplicación desarrollada para la participación ciudadana. Previo análisis cuantitativo, se desarrolló un marco conceptual clasificando iniciativas de acuerdo a nivel y área de participación, e identificando tecnologías y criterios de validación aplicados.

Las técnicas identificadas fueron abordadas en la Sección 2.3 del Capítulo 2, en particular son las siguientes: realidad aumentada, crowdsourcing, analítica de datos, minería de datos y de opinión, gamificación, sistemas de información geográfica, tecnologías semánticas, modelado de simulación, visualizaciones varias y herramientas para la visualización de argumentos. En el Apéndice A puede observarse la clasificación de todas las iniciativas recuperadas de la literatura.

En el nivel de e-Información, se pueden encontrar iniciativas que utilizan visualizaciones de diferentes tipos, realidad aumentada y componentes geo-referenciados, todas técnicas usadas para presentar información. En (KH14b) se analizaron los sitios web oficiales de los municipios eslovenos a fin de identificar qué herramientas de e-Participación se estaban utilizando para la democracia electrónica y qué herramientas electrónicas proporcionaba cada municipio a sus ciudadanos, siendo en su mayoría herramientas para compartir información. Por otro lado, los autores de (RBS⁺14) realizan un análisis cualitativo y cuantitativo de un proyecto de investigación llamado «ways2gether», que busca visualizar los cambios planeados en la ciudad usando realidad aumentada. De manera similar, la plataforma presentada en (LDD19) para permitir a los ciudadanos entender y observar la evolución espacio-temporal de sus ciudades. En (ARA11) también se utiliza la realidad aumentada en un prototipo para apoyar la participación pública en eventos de planificación urbana. Esta aplicación también permitía a los usuarios votar por sus diseños favoritos, y estos votos fueron recibidos por los responsables del evento urbanístico.

Las iniciativas en el nivel de e-Consulta permiten a los ciudadanos contribuir con su opinión; sin embargo, esto no implica que el gobierno efectivamente vaya a tener en cuenta estas contribuciones. Tal como en el nivel anterior, se identificaron iniciativas que muestran información acerca de la planificación urbana, promoviendo simultáneamente la comunicación con el ciudadano a través de encuestas y foros, como es el caso del proyecto «e-Community» de la ciudad de Narva (SS08). Otras herramientas pueden ser usadas para la planificación urbana, tal como información subjetiva y basada en emociones (PP17). La herramienta de crowdsourcing desarrollada por (PP17) recopila y visualiza información basada en emociones en mapas de la ciudad.

Varias iniciativas buscan involucrar al ciudadano en la identificación de los problemas de la ciudad. En (OW09) se realiza un estudio de caso longitudinal en profundidad sobre la plataforma «Taipei City Mayor's Mailbox», una fuente de información para (1) recopilar opiniones y quejas de los ciudadanos y (2) examinar el desempeño de cada de-

partamento en el servicio a los ciudadanos. Diferentes técnicas pueden ser utilizadas para analizar los reportes ciudadanos. Por ejemplo, en (DHS⁺17) se utilizan técnicas de *clustering* para obtener los temas en los reportes ciudadanos y para visualizar la intensidad o la acumulación de reportes en el área urbana. Los mismos autores utilizaron un modelo de clasificación para clasificar los reportes de manera automática (HDS17). En (FJ19) se presenta *FlashPoll*, una aplicación que permite involucrar a los residentes de un vecindario en los cambios de la ciudad, recopilando conocimientos y opiniones de los ciudadanos para tener en cuenta en la futura planificación urbana.

Algunos trabajos analizan el uso y los datos de redes sociales. En (AA15) se presenta cómo los municipios de Arabia Saudita utilizan Twitter para comunicarse con sus ciudadanos mediante el análisis del archivo histórico de todos los tweets municipales, buscando patrones de tweets, tipos y contenido de tweets, reacciones de los seguidores, entre otras estadísticas. Los autores de (SPSB17) exploran la gobernanza electrónica y las redes sociales en las evaluaciones de impacto y la posible contribución que pueden hacer a la participación pública. En (CLA⁺14) se presenta un enfoque novedoso de e-Participación basado en el «crowdsourcing pasivo», que puede definirse como la búsqueda de contenido y la extracción de conocimiento, ideas y opiniones sobre una política pública compartida por los ciudadanos sin ser alentada por los funcionarios públicos.

Otras formas de analizar las redes sociales es mediante el uso de tecnologías tales como «*raw analytics*» (análisis en bruto), minería de opinión y modelado de simulación, como se propone en (CL12) para la explotación centralizada de múltiples redes sociales para la elaboración participativa de políticas públicas. En (CGFL10) y (CL15) se describe una metodología para la explotación sistemática de redes sociales en combinación con técnicas de modelado de simulación que puede ser usado por organizaciones gubernamentales en procesos de formulación de políticas públicas. Otros trabajos usan minería de opinión para detectar y analizar automáticamente la postura del público hacia las decisiones gubernamentales (SCB⁺10) o para identificar una serie de medidas de tráfico implementables a corto plazo propuestas por los ciudadanos a través de canales digitales proporcionados por el gobierno (POB⁺12). Los datos recolectados de redes sociales también se pueden analizar usando un enfoque webométrico para identificar patrones de participación ciudadana y las relaciones entre ciudadanos, gobiernos y otras organizaciones involucradas en la formulación de políticas a través de las redes sociales (JPWP15).

Algunos artículos intentan incentivar la participación de los jóvenes en las políticas

públicas. La plataforma STEP busca involucrar a los jóvenes en los procesos de toma de decisiones relacionados a temas ambientales usando técnicas como gamificación y visualizaciones (VKS⁺17). Los autores de (TPP09) presentan los resultados de aplicar un framework basado en técnicas de inteligencia artificial sobre los datos de *mitmachen*, una plataforma destinada a motivar a los jóvenes a participar de debates políticos sobre temas de interés para el futuro de Austria.

Varias plataformas buscan iniciar un diálogo con el ciudadano, ya sea para establecer la comunicación entre ciudadanos y políticos a través de un foro de discusión en línea (SFS11) o para proponer y recoger nuevas ideas de soluciones a problemas específicos, como los relacionados con el cambio climático (RY15). Otras iniciativas están enfocadas en crear un canal de comunicación para involucrar a la ciudadanía en procesos de consulta y decisión. El framework propuesto en (Mur10) integra herramientas de conocimiento, consulta y voto electrónico en un portal regional que facilita los procesos de decisiones *bottom-up*. Los autores de (DCPS08) presentan una plataforma para superar los obstáculos de participación de los procesos locales de la Agenda 21 mediante el uso de un mapa de la ciudad con eventos de participación y diferentes herramientas deliberativas. En (QG18) se analiza la plataforma *Liberaopinion*, que sirve a los sindicatos docentes como herramienta de comunicación para realizar actividades como solicitudes y consultas electrónicas.

Se identificaron iniciativas que buscan involucrar a los ciudadanos en diálogos bidireccionales como parte de los procesos de gobernanza y formulación de políticas. En (SAAG18) se presenta una plataforma para la participación pública (*Liberopinion*), que consta de diferentes módulos, entre ellos uno para presupuesto participativo, gobernanza transparente y consultas públicas, entre otros. En (JWT11) se examina el proyecto *We-Gov*, que permite detectar, rastrear y extraer opiniones y debates sobre temas orientados a políticas. Otro ejemplo es *Puzzled by Policy*, una plataforma que usa funcionalidades como herramientas de visualización de argumentos e información analítica para mejor comprensión de los debates sobre temas de política (GR13). En (BM12) se presenta una herramienta para visualizar argumentos de propuestas de políticas como mapas de debate navegable. Otras plataformas están enfocadas en dar soporte a la e-Participación en procesos relacionados a la legislación. Los autores de (XL10) investigan el uso de un foro electrónico estructurado para apoyar la participación y la consulta electrónica en el proceso de formación de la legislación en los parlamentos. La plataforma de e-Participación descrita en (SFB⁺17; SFL⁺17) permite a los usuarios recuperar documentos prelegisla-

tivos, comentarlos y modificarlos, expresar sentimientos y valoraciones en las propuestas legislativas y en los comentarios, crear grupos, seguir otros usuarios, entre otros.

Las plataformas de e-Participación pueden ser usadas dentro de una universidad para mejorar la calidad de las decisiones mediante la participación, aumentando la transparencia y la legitimidad de las decisiones tomadas (PL18). La plataforma *MyUniversity* provee un conjunto de herramientas participativas configurables que permiten debatir, votar, usar cuestionarios en línea y distribuir información, entre otros (CÁP13). Este tipo de plataformas también pueden ser usadas en industrias, como es el caso del sistema de consulta en línea en el contexto de una autoridad reguladora croata (CVJ15), o en investigación y desarrollo, como la plataforma de crowdsourcing XPIR (VKIS16). Incluso pueden utilizarse para identificar soluciones a los problemas del cambio climático; por ejemplo, en (BGK⁺18) se presenta una plataforma para facilitar la colaboración de las partes interesadas en el análisis de las opciones de adaptación de la gestión del agua.

Algunos trabajos presentan un enfoque para la integración de diferentes modelos de participación. Los autores de (MP14) presentan un *integrated policy system* o sistema integrado de políticas, un enfoque multidisciplinario para la integración de modelos globales, individuales y de e-Participación de procesos de toma de decisiones. Este enfoque utiliza técnicas de simulación, análisis de redes sociales, análisis automático de texto y minería de opinión para facilitar la gobernanza electrónica integrada. En (WGJH13) se presenta algunas de las herramientas de participación móvil aún en desarrollo, como encuesta móvil, consulta móvil y navegador de privacidad, junto a la arquitectura para la integración de estas herramientas de participación móvil.

El nivel de e-Colaboración permite a los ciudadanos participar activamente en el desarrollo de alternativas y en la identificación de las soluciones deseadas. Muchas plataformas permiten a los ciudadanos reportar problemas relacionados a diferentes temas de la ciudad, como puede ser un proyecto para mapear basurales ilegales (KH14a; BJB⁺18), un sistema para reportar quejas del sistema de transporte público (FPB⁺17) o enviar solicitudes (propuestas o sugerencias) relacionados a temas urbanos (VT18). Otros sistemas permiten a los ciudadanos mantenerse informados de los últimos anuncios y eventos, proporcionar retroalimentación sobre los proyectos y políticas impuestas, informar inquietudes o problemas (BFF15) o co-construir información de accesibilidad de la ciudad (EA). Otro ejemplo de plataforma de e-Colaboración es un sistema de peticiones electrónicas, como el caso de estudio de Sutharyakeralam, donde los ciudadanos pueden registrarse y enviar sus peti-

ciones y luego rastrear su estado (SWA⁺12). Las organizaciones sociales también generan iniciativas de e-Participación para promover campañas sociales, culturales y políticas, como *Nossas Cidades* (dCPdSA19).

Muchas plataformas en este nivel usan componentes de juegos para incentivar la participación ciudadana, en especial en la planificación urbana. El autor de (Thi16b) revisa proyectos académicos recientes relacionados con herramientas de participación gamificadas e informa sobre los resultados preliminares, como *Love your city!* y *Täsä*. Otro ejemplo es la plataforma *Community Circle* que permite a los ciudadanos y funcionarios de la ciudad presentar temas para el debate público mediante contribuciones (LBE⁺14). Algunos aspectos de juego están relacionados al área de impacto de las contribuciones, su vida útil, la participación en las misiones creadas por los funcionarios públicos, entre otros (TLSG15; TL15). Plataformas similares se centran en involucrar a los ciudadanos en la creación y el intercambio de propuestas con otros ciudadanos y con la administración pública (BFR16) o en informar problemas y hacer contribuciones, así como facilitar comentarios sobre un plan o tema de administración específico (TRR16). Un enfoque diferente es propuesto en (RFT18), que involucra a los jóvenes en el proceso de planificación urbana utilizando el juego *Minecraft*, modelado de simulación y pensamiento de diseño. Otros artículos exploran si la incorporación de elementos de juego afecta el nivel y la calidad de la participación y en qué manera (TEB17; Thi16a), y si las variables demográficas afectan la participación y la motivación cuando se utiliza una aplicación gamificada (TRRB19). El autor de (Ber12) utiliza el modelo YUPTA para la evaluación de diferentes proyectos de e-Participación, como el sitio interactivo para los ciudadanos de Breda o el sitio web «*De Amstel Verandert*», donde los ciudadanos pueden proponer sus ideas acerca del futuro del área del río Amstel. El autor presenta también el proyecto *Plan-o-Mat*, una herramienta en línea donde los ciudadanos pueden hacer planes con medios simples, ayudando a descubrir las preferencias de los usuarios. Los elementos de juego pueden ser combinados con las redes sociales para reducir las barreras a la participación en la planificación urbana (JH18). En (KLLS14) se presenta cómo se pueden desarrollar diferentes tipos de TIC para la planificación urbana y la toma de decisiones, como presentar los planes de desarrollo urbano con la ayuda de escenarios 3D u obtener información sobre la distribución de la población y los patrones de movilidad.

Muchas plataformas se enfocan en facilitar el debate respecto a políticas públicas. En (POB18) se presenta una infraestructura que actúa como un puente hacia las deliberacio-

nes políticas de e-Participación lideradas por el gobierno y por los ciudadanos, aquellas que provienen de la e-Participación tradicional y las deliberaciones lideradas por ciudadanos a través de las redes sociales. Otro ejemplo es la iniciativa ciudadana europea, que permite a los ciudadanos participar en el desarrollo de las políticas de la Unión Europea pidiendo a la Comisión Europea que adopte una propuesta legislativa (SG14). En (Too19) se estudia una plataforma para involucrar a los ciudadanos en consultas sobre proyectos de legislación y políticas, así como obtener propuestas de política por parte de los ciudadanos. También hay plataformas para involucrar a los ciudadanos en el proceso de sugerir, debatir y/o decidir la asignación de una porción del presupuesto público administrado por el gobierno (ZOC18). También se desarrollan herramientas para facilitar la comprensión y el debate, como una herramienta de visualización de argumentos para participar en debates y explorarlos como si fuera un mapa (TDPT11), o como la plataforma descrita en (AKLC18) para apoyar la consulta y argumentación entre expertos sobre problemas sociales y políticas públicas. Otros trabajos presentan iniciativas relacionadas a procesos democráticos. En (OAOM16) se presenta un framework para la democracia electrónica después de un análisis cualitativo de los *frameworks*, modelos y estrategias de democracia electrónica existentes en la literatura. Los autores de (POB16) presentan una ontología para e-Participación que cubre las facetas centrales de la e-Participación: como un proceso económico, como una iniciativa y como un sistema socio-técnico.

Varios artículos científicos tratan sobre plataformas diseñadas para involucrar a los ciudadanos en la formulación de políticas. Algunos tienen como objetivo informar, involucrar y permitir a los ciudadanos contribuir a los procesos de toma de decisiones gubernamentales y de políticas (CFE13), proporcionando herramientas como un servicio de mapas de políticas basado en la ubicación, minería de opinión y tecnologías semánticas (AM12). En (ENV14) se presenta una plataforma deliberativa para la elaboración de reglas que permite aprender y proporcionar información sobre las discusiones sobre políticas gubernamentales. Los autores de (CME⁺13) presentan un framework para identificar y contrastar automáticamente patrones de pensamiento colectivo y para formalizar argumentos para la minería de opinión en Twitter (CMG14). Las redes sociales también se pueden utilizar con otros fines. En (SCLD14) se presenta un framework para recopilar conocimientos, ideas y opiniones de los ciudadanos y aplicar prácticas de crowdsourcing en el sector público. Los autores de (SA15) presentan un modelo conceptual para realizar reglamentación electrónica en redes sociales y estudiar las variables que pueden afectarlo, como las interfaces de usuario, la privacidad, la seguridad y los grupos de influencia.

3.3. Uso de tecnologías en iniciativas de participación ciudadana

En la Figura 3.1 se presenta una línea de tiempo que muestra el uso de las tecnologías a lo largo de los años en iniciativas de participación ciudadana presentes en la literatura. Esta línea de tiempo abarca desde la primera plataforma identificada, es decir, desde el 2007 al 2019. En la figura podemos observar en qué año las tecnologías (descritas en la Sección 2.3.2) comenzaron a utilizarse y cómo fue su evolución respecto a su uso. Cada fila de símbolos representa una iniciativa diferente, y cada símbolo representa una tecnología diferente utilizada en esa iniciativa. Cada tecnología tiene asociado un símbolo, como se indica en la leyenda de la figura.

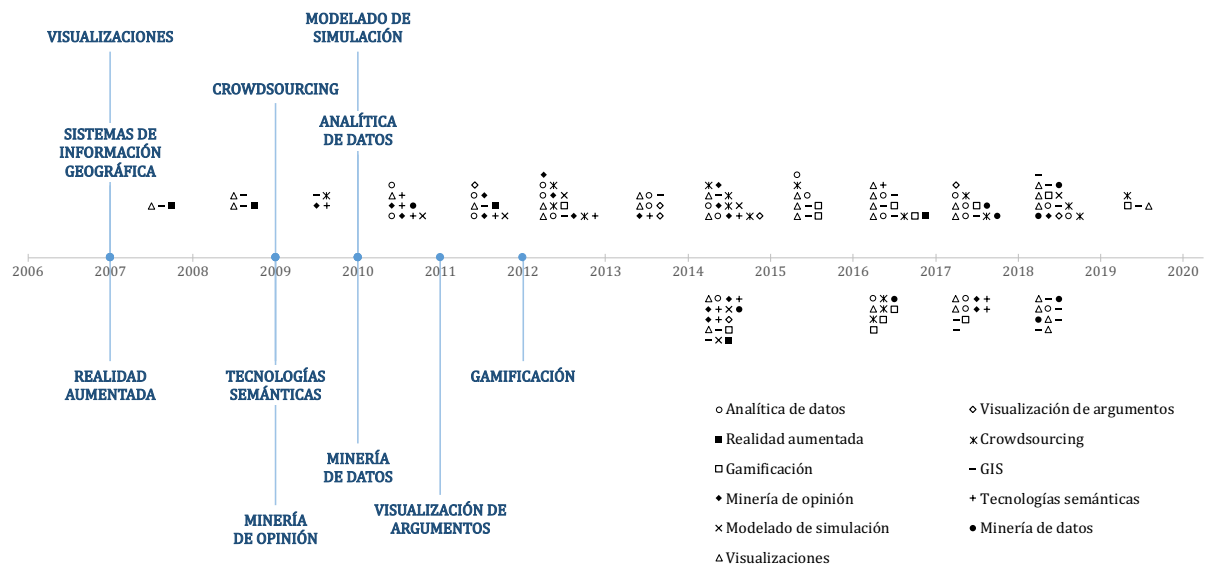


Figura 3.1: Línea de tiempo de la evolución en el uso de las tecnologías.

De la figura se puede derivar que los años más productivos en términos de iniciativas (publicadas en trabajos de investigación) fueron el 2014 y el 2018 (cada año con nueve iniciativas), seguidos por los años 2016 y 2017 (que cuentan con ocho iniciativas cada uno). También se puede analizar el uso de técnicas por año. Según nuestra búsqueda, las técnicas de analítica de datos se utilizaron por primera vez en 2010 y luego su uso se volvió recurrente. Algo similar ocurre con las GIS y las visualizaciones (con su primera aparición en 2007), ambas técnicas son muy populares y útiles en todos los niveles y áreas. Por el contrario, la realidad aumentada apenas se utilizó desde su primer uso en 2007 (cero o un

uso por año). La evolución en el uso de las herramientas de visualización de argumentos y el modelado de simulación es similar.

Según nuestra investigación, la minería de datos se utilizó por primera vez en 2010. Luego, se usó de forma esporádica hasta el 2016, cuando comenzó a tener un uso más constante. La minería de opinión y las tecnologías semánticas tienen un crecimiento en su uso similar: ambas técnicas se utilizaron por primera vez en 2009, tuvieron un uso regular con un pico en 2014 y luego tuvieron una disminución en su uso. El uso crowdsourcing y gamificación también es similar, con su primera aparición en 2012. El crowdsourcing se utilizó ampliamente solo luego de su primer pico en 2014. El uso de la gamificación creció con el tiempo, con un pico de uso en el 2016, para luego disminuir. El aumento en el uso de estas técnicas puede estar asociado con el aumento en la creación de iniciativas que buscan alentar a los ciudadanos a participar activamente en la formulación de políticas e identificación de problemas en la ciudad, entre otros.

3.4. Tipo de validaciones

Para comprender cómo se evalúan las técnicas, identificamos los métodos utilizados para validar los resultados presentados en cada uno de los trabajos de investigación seleccionados. En particular, reconocimos iniciativas validadas mediante la realización de enfoques cualitativos, cuantitativos y ambos, como se analiza a continuación.

3.4.1. Validación cualitativa

El análisis cualitativo utiliza un juicio subjetivo basado en información no cuantificable. Como ejemplo, en (RY15) se realizaron dos evaluaciones generales en experiencias de crowdsourcing: (1) del proceso en cuanto a transparencia, participación y colaboración, y (2) del resultado, donde los componentes clave son la efectividad del producto final y el impacto de los insumos participativos en el trabajo del gobierno. En (SG14) se recopilaron datos de estudios de documentos, observaciones y entrevistas, y luego se llevó a cabo un análisis desde tres perspectivas diferentes: explorando la interacción entre las diferentes dimensiones que rodean el desarrollo de la e-Participación, identificando los factores para una implementación más exitosa, e investigando e interpretando los fundamentos que subyacen a los pasos y decisiones cruciales de las instituciones.

Otros enfoques están relacionados con realizar mejoras en la aplicación utilizando los comentarios de los usuarios (DCPS08; BM12). En (BM12) se adoptó un enfoque de taller basado en escenarios en el que se utilizan breves clips narrativos para cubrir cómo se pretende utilizar el prototipo, y también se estipuló la discusión sobre la viabilidad y conveniencia de cualquier cambio en el prototipo. En (AM12) se comparan las funcionalidades de UbiPOL con otras herramientas e iniciativas de e-Participación que brindan los mismos servicios. Luego, realiza un análisis FODA del sistema. En (Ber12) se utiliza YUPTA para la evaluación de diferentes proyectos de e-Participación. YUPTA es un modelo de comunicación que comprende cuatro dimensiones (tiempo, lugar, acción y relación) que definen cómo interactúan las personas.

3.4.2. Validación cuantitativa

La mayoría de las iniciativas clasificadas aplican la evaluación estadística de los datos recopilados durante el experimento. Una forma de recopilar datos es a través de preguntas de encuestas que pueden estar relacionadas con información demográfica, interés en la política y sobre la aplicación en sí, como los cuestionarios distribuidos en (ARA11) y (TRR16). En (BFF15) se usan cuestionarios para identificar: (1) el nivel de participación de la comunidad en el monitoreo del servicio público, (2) el nivel de percepción de la comunidad sobre el uso de la tecnología, y (3) el nivel de aceptación de la comunidad hacia el sitio web (referido a su funcionalidad, usabilidad, confiabilidad y desempeño).

Otra forma de recopilar datos es a través del uso de la plataforma, monitoreando las actividades en la plataforma y almacenando información estadística relacionada con el acceso (QG18). En (VKIS16) estos datos recopilados se utilizan para construir un modelo de clasificación automática relacionada con el uso del sistema. Independientemente de la falta de un framework específico, en (AA15) se evaluó «cómo» el gobierno usa Twitter a través de datos estadísticos (como tweets por día, porcentaje de respuestas, número promedio de enlaces y número medio de retweets), análisis de datos (por ejemplo, coeficientes de correlación o patrones de uso) e identificación de diferentes tipos de tweets. En (SCB⁺10) se presenta una validación que está fuertemente relacionada con el framework propuesto. A partir de una configuración experimental, se recopila y procesa opiniones de ciudadanos reales, se extraen distintas fases de opinión y finalmente se compara su desempeño mediante la capacitación y ejecución de diferentes clasificadores.

3.4.3. Validación cuantitativa y cualitativa

Las iniciativas y tecnologías aplicadas se evalúan posteriormente mediante análisis cuantitativo y cualitativo. Por ejemplo, en (XL10) se realizó un análisis de un árbol de discusión que incluye el cálculo de métricas relacionadas con las publicaciones del árbol. En (BFR16) se realiza un estudio de usuario para probar la aplicación con y sin los elementos de gamificación. Otras iniciativas crearon cuestionarios para partes interesadas específicas, como académicos, consultores, legisladores (TDPT11) o expertos en sistemas de información y legislación (SA15).

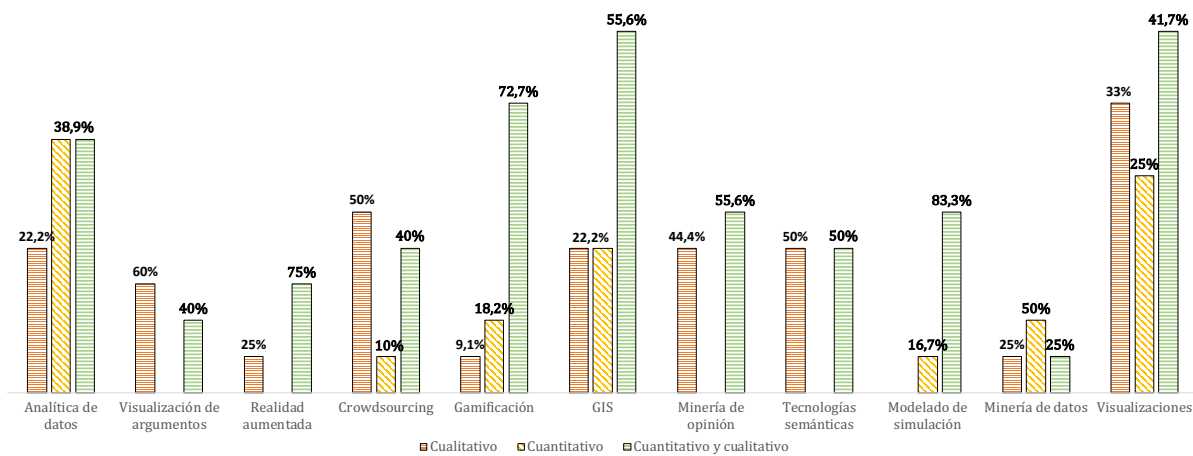


Figura 3.2: Proporción del tipo de validación con respecto a las tecnologías utilizadas.

3.4.4. Validación y tecnologías

La Figura 3.2 muestra el tipo de validación realizada en función de las tecnologías utilizadas. Podemos identificar fácilmente los tipos de validación más populares, que es cuando se aplican ambas técnicas (cualitativa y cuantitativa) simultáneamente. También podemos observar que algunas tecnologías comparten tipos de validaciones en proporciones similares, como es el caso de la minería de opinión y las tecnologías semánticas. Podemos concluir que algunas técnicas están más probadas que otras, pero esto probablemente se deba a que son más utilizadas.



Figura 3.3: El gráfico describe la relación entre los niveles y las áreas de participación.

3.5. Evaluación del estado del arte

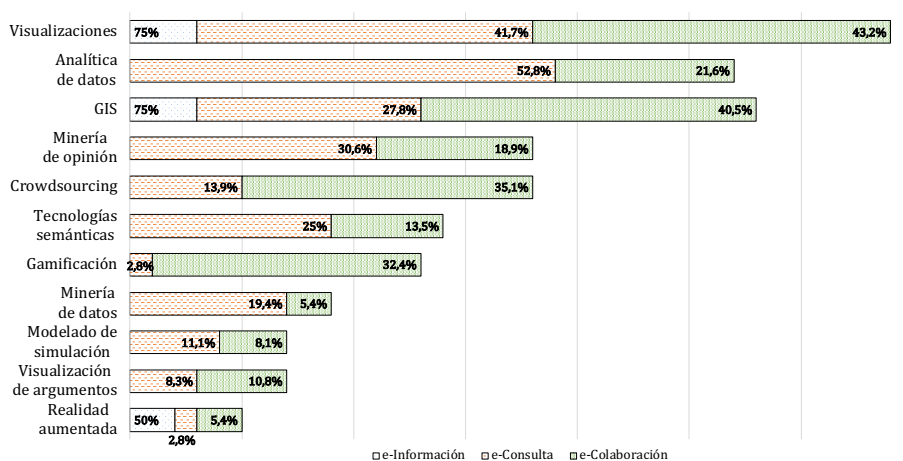
En esta sección se realizará un análisis de los trabajos presentados en la Sección 3.2. En particular, se explorarán las relaciones entre nivel y área de participación, y el uso de las tecnologías en diferentes niveles, áreas y dependiendo de quién las inicia. Este análisis también busca responder en parte la PI2, definida con anterioridad en la Sección 1.2.

En la Figura 3.3 se puede observar por cada nivel de participación cuáles son las áreas más populares. En particular, para el nivel e-Información, se identificaron cuatro categorías, de las cuales tres corresponden al área de planificación territorial. La existencia de estas plataformas parece indicar el deseo de los gobiernos de comunicarse y mantener informados a los ciudadanos sobre los cambios que se están produciendo en la ciudad. En el nivel e-Consulta, además del área consulta, otras áreas populares son la deliberación y votación; ambos permiten el análisis de discusiones y opinión pública sobre temas de interés. Estas áreas en combinación con el uso de las técnicas adecuadas permiten a las partes interesadas recopilar información valiosa para la toma de decisiones. Finalmente, en

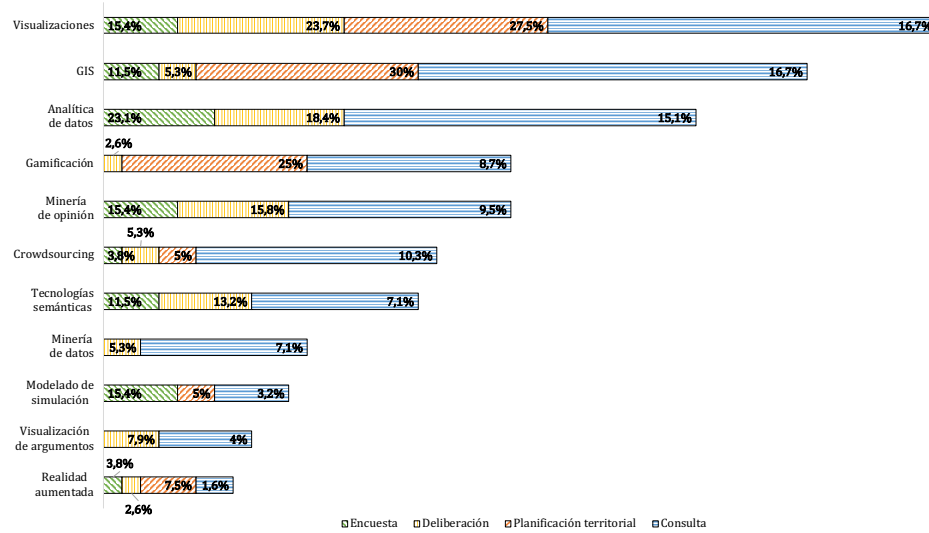
el nivel e-Colaboración, las áreas más populares son la consulta, la planificación territorial y la deliberación. Además, se puede observar que la proporción de iniciativas en las áreas de construcción comunitaria y campaña electoral aumentó en comparación al nivel anterior.

Se puede profundizar el análisis examinando qué técnicas se utilizan generalmente en los diferentes niveles (ver Figura 3.4a) y en las áreas más populares (ver Figura 3.4b). La mayoría de las técnicas, especialmente las visualizaciones, se pueden aplicar a cualquier nivel o área de participación. Aún así, algunos de ellos resultan más útiles en ciertas áreas. En particular, la minería de opinión, las tecnologías semánticas y la analítica de datos son las técnicas más comunes en el área de deliberación, ya que pueden utilizarse para analizar y extraer información de textos, como discusiones o debates. En el área de planificación territorial, las técnicas más utilizadas son las GIS, la gamificación, la realidad aumentada y el crowdsourcing, que pueden ser útiles para representar información y para involucrar a la ciudadanía en las iniciativas. Las técnicas más comunes para el área de encuestas son la analítica de datos, el modelado de simulación, la minería de opinión y las tecnologías semánticas, todas técnicas que ayudan a medir el sentimiento y la opinión pública.

Se puede observar una relación más fuerte entre las técnicas utilizadas y el nivel de participación. Las técnicas en e-Información son las visualizaciones, la realidad aumentada y las GIS, dada su capacidad para mejorar la presentación de información. Las técnicas aplicadas en el nivel de e-Consulta están orientadas a la evaluación de opiniones y preferencias, tales como la minería de opinión, las tecnologías semánticas y la analítica de datos. Por otro lado, técnicas como GIS, crowdsourcing y gamificación se utilizan típicamente en el área de e-Colaboración, dado que implican la interacción con usuarios y tienen



(a)



b

Figura 3.4: Los gráficos representan el uso de técnicas en cada (a) nivel de participación y (b) área de participación.

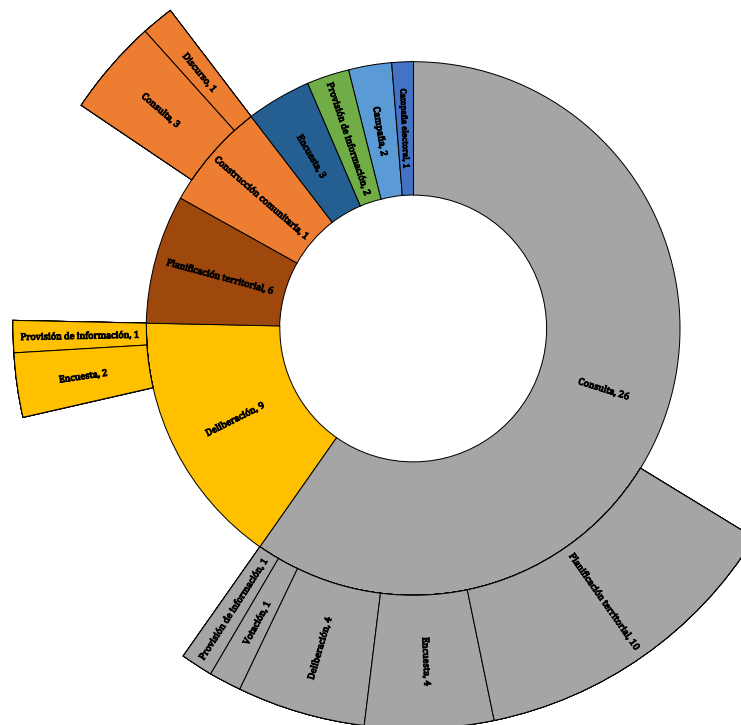
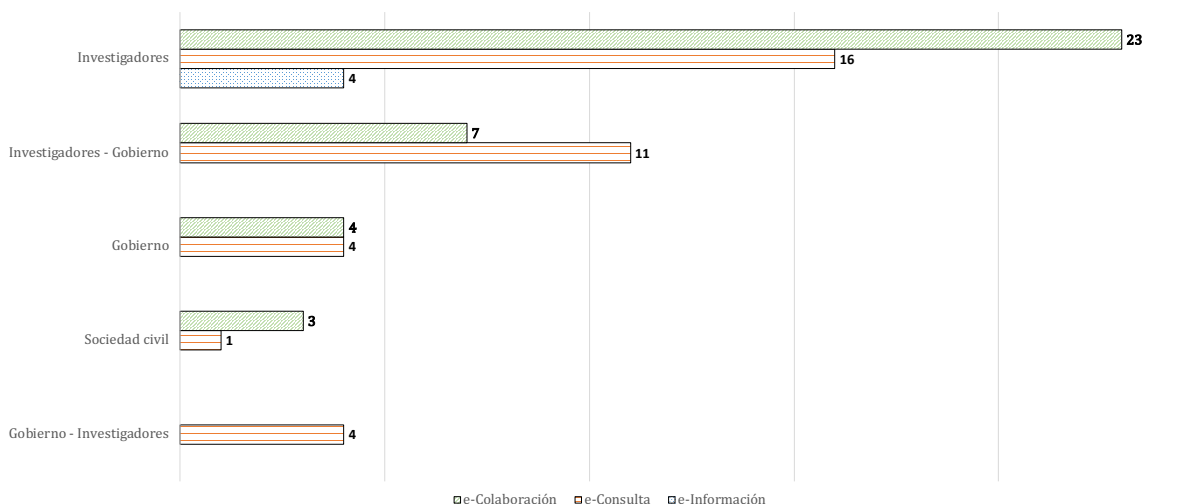


Figura 3.5: Relación entre las áreas de participación.

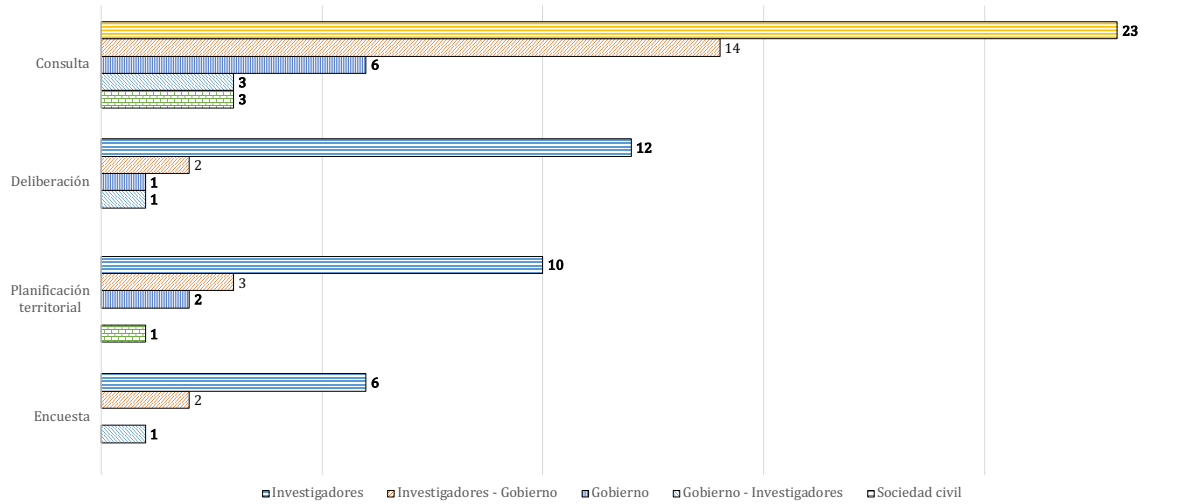
como objetivo fomentar la participación ciudadana. Esto podría implicar que la correcta selección de técnicas es importante para obtener la participación activa de la ciudadanía.

La Figura 3.5 representa la relación entre las áreas de participación: para cada área (primaria), se indica qué y en qué grado otras áreas la complementan. Así, se puede identificar qué áreas tienen más probabilidades de definir la cobertura de una iniciativa en conjunto. El área de consulta puede combinarse con la mayoría de las áreas, lo que puede indicar que es ampliamente utilizada ya que permite una comunicación e interacción más directa con los ciudadanos. En particular, se la combina mucho con la planificación territorial, sobre todo en el nivel de e-Colaboración. En el caso del área de deliberación, suele combinarse con el área de votación.

En la Figura 3.6a se puede observar la relación entre el origen de una iniciativa y los niveles de participación. Podemos concluir que los investigadores generalmente desarrollan iniciativas en el nivel de e-Colaboración, mientras que las iniciativas dirigidas por el gobierno todavía se desarrollan en el área de e-Consulta, aunque se observa que van avanzando hacia el siguiente nivel. La Figura 3.6b presenta la relación entre el origen de la iniciativa y las áreas de participación. Las iniciativas dirigidas por el gobierno generalmente se enfocan en el área de consulta. En colaboración con investigadores, también se desarrollan iniciativas en las áreas de deliberación y votación. Por otro lado, las plataformas propuestas por investigadores abarcan todas las áreas. Sin embargo, aquellas en colaboración con gobiernos, se centra en las áreas de planificación urbana y consultas. Estas mismas áreas son de interés para las iniciativas lideradas por sociedades civiles.



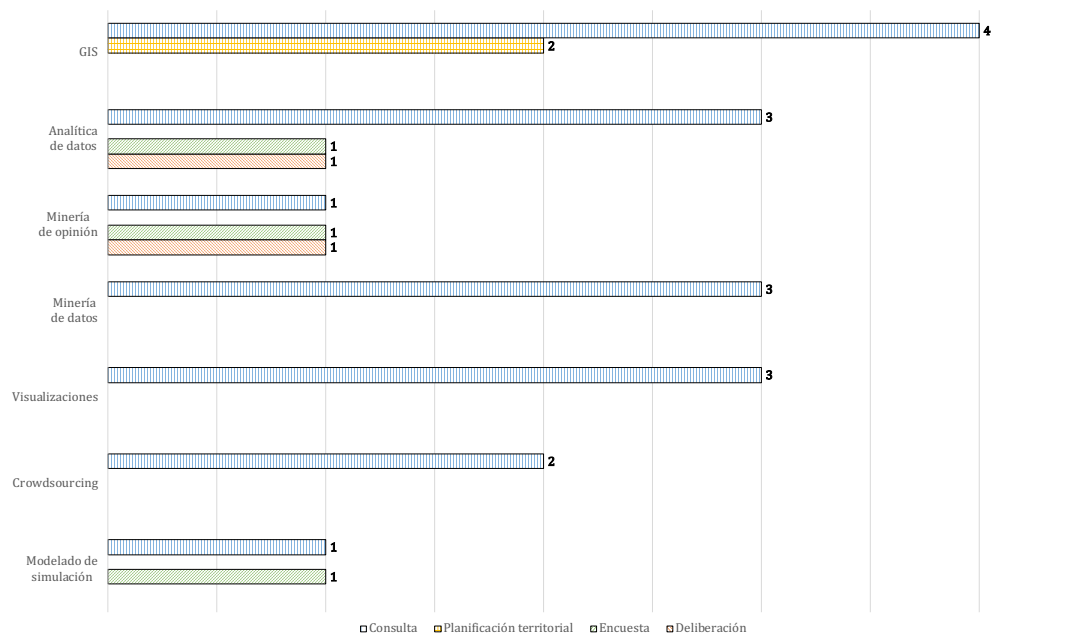
(a)



(b)

Figura 3.6: Los gráficos muestra el número de iniciativas por origen en cada (a) nivel de participación y (b) área de participación.

La Figura 3.7a muestra la distribución de las tecnologías utilizadas en las áreas de participación de iniciativas gubernamentales. Todas las iniciativas originalmente propuestas por el gobierno (y también las dirigidas por el gobierno con participación de investigadores) pertenecen mayormente al nivel de e-Consulta y las áreas abordadas son la planificación



(a)

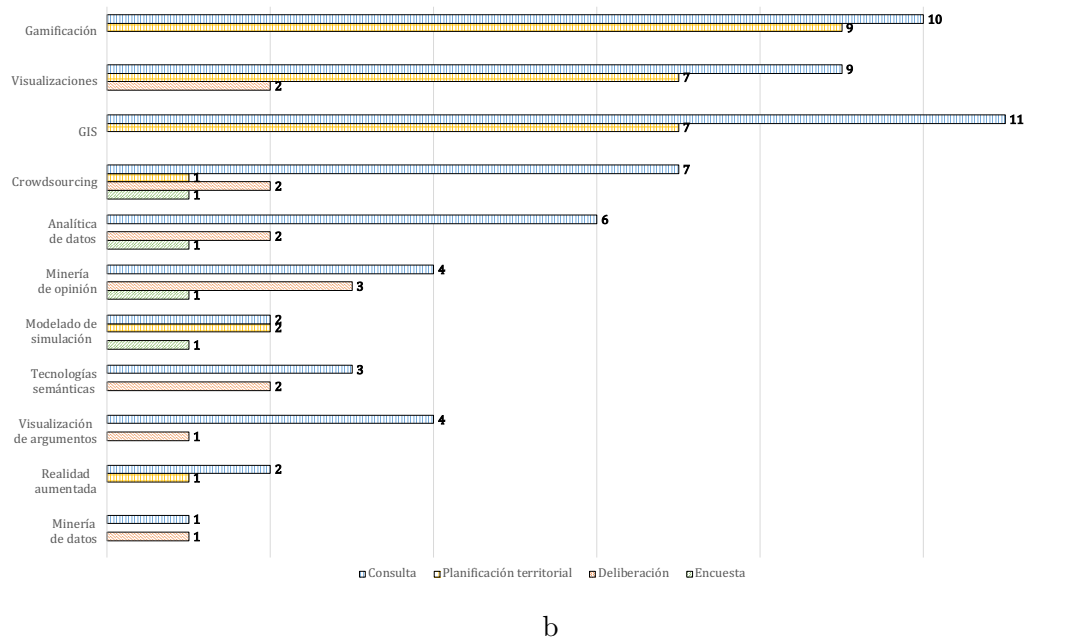


Figura 3.7: Los gráficos muestra las tecnologías utilizadas en cada nivel de participación por las iniciativas (a) dirigidas por el gobierno –con o sin participación de investigadores– y (b) lideradas por la investigación –con o sin ayuda gubernamental– en el nivel de e-Colaboración.

territorial, la votación, la deliberación y la consulta. Se destaca el uso de la minería de opinión y la analítica de datos en el área de deliberación, el de estas dos técnicas junto con el modelado de simulación para el área de votación, y el uso de GIS para el área de planificación territorial. De manera similar, la Figura 3.7b muestra la distribución de tecnologías utilizadas en las áreas de participación más populares de las iniciativas lideradas por la investigación (o lideradas por la investigación con ayuda gubernamental) en el nivel de e-Colaboración. En las áreas de planificación territorial y consulta, se utilizan mayormente las visualizaciones, las GIS y la gamificación. El crowdsourcing es una de las técnicas más utilizadas en las áreas de votación y deliberación, posiblemente porque el nivel de e-Colaboración busca involucrar a los ciudadanos en la propuesta de políticas. En el área de deliberación también se hace uso de minería de opinión, tecnologías semánticas y la analítica de datos; esto puede estar relacionado al interés de los investigadores en desarrollar y aplicar dichas técnicas.

3.6. Resumen

En este capítulo se realiza un análisis de trabajos relacionados. Luego, el análisis se enfocó en la aplicación de técnicas informáticas en iniciativas de e-Participación, recuperando diferentes artículos científicos que analizan herramientas o plataformas de participación ciudadana. Estas iniciativas fueron clasificadas de acuerdo a su nivel y área de participación, y se identificaron las tecnologías y criterios de validación utilizados. Finalmente, se presentó un análisis sobre dichas iniciativas, enfocándose en particular en las relaciones existentes entre niveles y áreas de participación, y en el uso de las tecnologías en estos niveles, áreas y dependiendo del origen de la iniciativa.

Capítulo 4

Herramientas informáticas para el análisis de comunicaciones ciudadanas

En este capítulo se presenta una plataforma para el agrupamiento y exploración visual de elementos geocalizados, desarrollado especialmente para analizar reportes ciudadanos. La Sección 4.1 presenta la motivación. La metodología de desarrollo se indica en la Sección 4.2, mientras que en la Sección 4.3 se muestran las diferentes funcionalidades de la herramienta. Las secciones 4.4 y 4.5 presentan su diseño, implementación y testeo. La validación de la herramienta se realiza mediante dos casos de estudio, que se presentan en las secciones 4.6 y 4.7. Finalmente, las contribuciones y las limitaciones de la plataforma desarrollada se discuten en la Sección 4.8.

4.1. Motivación

Las PQR de los ciudadanos proporcionan una valiosa fuente de información sobre los problemas relacionados con la infraestructura y los servicios de los municipios. La aplicación de las TIC en la administración pública para recopilar y procesar las PQR ciudadanas proporciona a las partes interesadas (ciudadanos y funcionarios gubernamentales, entre otros) canales electrónicos adecuados para la eficiente comunicación y posterior coordinación de la solución (OW13; Hon13; SE15). En el contexto particular de la gestión

inteligente de una ciudad, se han desarrollado una serie de plataformas para escuchar y dar respuesta a las PQR ciudadanas (FTA12; DB12; SKS15). Sin embargo, las plataformas existentes solo brindan una solución parcial al problema de analizar de manera efectiva y eficiente las grandes cantidades de datos que generan diariamente sus usuarios. En algunos casos, las PQR ciudadanas corresponden a cuestiones aisladas, mientras que en otros se asocian a problemas mayores. Las tendencias relacionadas con ciertas temáticas podrían corresponder a problemas persistentes que impactan áreas geográficas específicas o tipos particulares de servicios, que las autoridades relevantes deben identificar y resolver priorizando las demandas más urgentes.

Este estudio está impulsado por la pregunta de investigación definida previamente en la Sección 1.2: **¿Qué componentes y técnicas pueden ser parte de una infraestructura de software que haga uso de información volcada en plataformas digitales para facilitar la toma de decisiones por parte de funcionarios públicos?** Se busca el diseño e implementación de una herramienta que incluya un conjunto de algoritmos de minería de datos y técnicas de visualización apropiadas para los datos relacionados a PQR ciudadanos, teniendo en cuenta su condición de datos geolocalizados.

En las secciones posteriores, se presenta una herramienta de software para complementar el análisis de PQR ciudadanas mediante una serie de funcionalidades interactivas para el análisis visual y la exploración de dichas solicitudes. Estas PQR ciudadanas pueden provenir de plataformas privadas o estatales, tal como se mostrará más adelante en los casos de estudio. Las técnicas utilizadas y las funcionalidades de la herramienta se presentarán en secciones posteriores.

4.2. Metodología de desarrollo

La implementación de la herramienta siguió una metodología de desarrollo evolutiva. Esta metodología implica la construcción de una versión inicial que se irá refinando de acuerdo a nuevos requerimientos por parte del usuario, donde las fases de especificación, desarrollo y validación se entrelazan. Se eligió esta metodología porque permite de forma más natural la exploración de nuevos requerimientos, donde la especificación del sistema puede desarrollarse de forma creciente e interactiva, logrando un mejor entendimiento del sistema y siendo más efectiva para nuestro objetivo que el modelo tradicional de cascada.

Dada la naturaleza exploratoria del desarrollo, se definieron diferentes funcionalidades y las mismas se fueron implementando en los diversos ciclos. Las siguientes secciones explican las características más importantes del desarrollo.

4.3. Requerimientos

Las principales funcionalidades de la herramienta se ven reflejadas en el diagrama de caso de uso presentado en la Figura 4.1. Un usuario de la herramienta será el administrador de sistema, encargado de *Importar los datos* que se utilizarán luego. El otro usuario de la herramienta será un oficial de gobierno, es decir un funcionario o empleado público que tomará decisiones sobre cómo afrontar las problemáticas presentes en la ciudad. El usuario puede utilizar las tres principales funcionalidades del sistema: *Visualizar datos estadísticos*, *Visualizar mapa de calor* y *Visualizar mapa de agrupaciones*.

La funcionalidad de datos estadísticos le permite al usuario observar información sobre los problemas en los diferentes servicios de la ciudad, permitiendo tener una visión de las dimensiones de los problemas del municipio de una manera clara e inmediata. El mapa de calor le permite distinguir las concentraciones de reportes en diferentes zonas geográficas, de manera que le será más fácil reconocer las áreas más afectadas. Finalmente, el mapa de agrupaciones le ofrece la posibilidad de observar agrupaciones de reclamos y sus límites geográficos en el mapa, permitiendo la definición de estrategias para atender mejor las demandas de los ciudadanos teniendo cuenta el tipo de servicio y su cercanía geográfica. Estas funcionalidades se explicarán en detalle en las siguientes subsecciones.

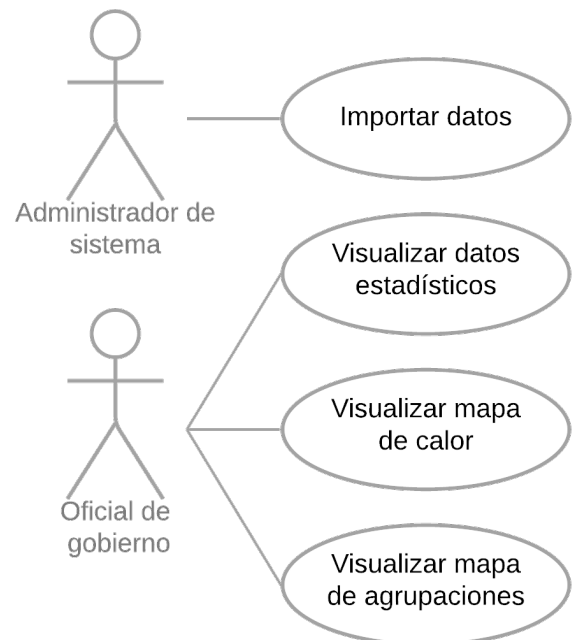


Figura 4.1: Diagrama de caso de uso.

4.3.1. Importar datos

Esta funcionalidad ofrece una sencilla interfaz al administrador del sistema para que importe los datos de reclamos ciudadanos a la base de datos en formato CSV o XLSX. Estos datos serán posteriormente accedidos por el servidor en base a los pedidos de usuario.

4.3.2. Visualizar datos estadísticos

Esta funcionalidad provista por la herramienta permite visualizar una serie de datos estadísticos calculados sobre el conjunto total de reclamos, que se encuentran representados visualmente por medio de un gráfico de proyección solar. Este gráfico permite presentar información organizada por jerarquía, donde el centro del gráfico corresponde al total de servicios (y de reportes provistos) y las secciones alrededor corresponden a los diferentes servicios y categorías. Para determinar el tamaño de cada sección, se le asigna un valor a cada categoría de cada servicio que se corresponde con la cantidad de reportes asociados a la categoría en cuestión. De este modo, los diferentes valores de cada categoría determinan el valor del servicio que conforman. A su vez, los valores calculados por los servicios conforman el valor global del total de los servicios (en el centro del gráfico).

Junto al gráfico hay una sección de información en donde pueden observarse los porcentajes de reportes generados sobre cada uno de los servicios como una lista de ítems. En el caso de un servicio en particular, se mostrará una lista con sus categorías, su proporción dentro de ese servicio y su proporción respecto al total de reportes. Estas proporciones se calculan utilizando la cantidad de reportes realizados para cada categoría de cada servicio. Además, tanto la sección de información (abajo del título del servicio o categoría) como el gráfico (al mantener el mouse sobre una determinada sección) muestran la cantidad de reportes generados para el total de los servicios, un servicio o una categoría determinada. Se pueden identificar los servicios y categorías en el gráfico pasando el mouse sobre las secciones. Una característica especial de esta implementación de gráfico de proyección solar es que permite realizar zoom en cualquier arco o subsección que se cliquee, pudiendo disminuir el zoom al hacer clic en el centro del gráfico.

Otra característica de la herramienta es que permite identificar las secciones del gráfico que se corresponden con los elementos de la lista en la sección de información. Al pasar el puntero por encima de algún ítem de la lista, la sección correspondiente en el gráfico se colorea de negro. Además, al hacer clic sobre el ítem, se realizará *zoom in* tanto en el

gráfico sobre la sección elegida, así como también en la información asociada a esa misma porción. Para acompañar esta navegación por fuera del gráfico, se tienen dos botones que permiten volver a la sección anterior o directamente a la totalidad de los servicios.

Para esta funcionalidad, el cliente realiza un pedido al servidor, que devolverá un archivo JSON con la información correspondiente. Para que sea de utilidad a funcionarios públicos, es necesario que este archivo JSON sea generado de manera regular, de modo que esté lo más actualizado posible a fin de que se visualicen solo las problemáticas que no han sido resueltas y su incidencia en cada categoría, servicio o sobre el total.

4.3.3. Visualizar mapa de calor

El mapa de calor permite visualizar la intensidad de las solicitudes y quejas en la ciudad, ofreciendo una valiosa ayuda para reconocer las áreas más afectadas. Para la creación y representación visual del mapa de la ciudad, se utiliza una directiva provista por Mapbox¹, indicando las coordenadas geográficas centrales de la ciudad, el nivel de zoom inicial, y el nivel mínimo y máximo de zoom permitidos. Para el funcionamiento y la correcta visualización del mapa de calor, se debe asignar un *source* al mapa y establecer la opción de clúster como *true*. El contenido de este *source* será el GeoJSON que contenga la información geográfica requerida por Mapbox. La escala de valores para los grados de calor utilizada para el mapa de calor está definida en un archivo JSON (llamado *layers*), donde se define un arreglo que contiene pares detallando la cantidad de puntos requeridos para formar un grado y el color asignado a ese grado. Este archivo es utilizado por la función *heatmap* provista por Mapbox. En el mapa se incluye una sección de referencia para los diferentes gradientes, utilizándose colores más cálidos a medida que aumenta la concentración de reportes en una misma zona del mapa. De este modo, los colores del mapa varían en un rango desde verde (ligeramente afectado), pasando por amarillo (moderadamente afectado), hasta llegar al rojo (muy afectado). Utilizando directivas provistas por Mapbox, se agregan nuevas capas al mapa, asociadas al *source* asignado y en base al arreglo de *layers* definido con anterioridad. De este modo, quedan vinculados los clústeres generados por Mapbox con las *layers*, y, al agregarse las capas al mapa, se puede apreciar el mapa de calor según la escala utilizada.

¹ Mapbox - <https://www.mapbox.com/>

La herramienta desarrollada ofrece opciones de *zoom in* y *zoom out*, haciendo posible visualizar la ciudad en su totalidad o enfocarse en áreas específicas. Además, al llegar al máximo nivel de *zoom in* podrán verse los reportes individuales como puntos de color azul. Adicionalmente, es posible visualizar un mapa de calor para el conjunto total de servicios, o seleccionar y ver el mapa de calor para un servicio específico. Para esto, al cargar la página, se recuperan los GeoJSON generados para cada servicio, de modo que al cambiar de opción solo reste borrar las capas asociadas al *source* utilizado hasta el momento, y agregar al mapa las capas que corresponden a la opción elegida.

4.3.4. Visualizar mapa de agrupaciones

Esta funcionalidad permite visualizar agrupaciones de PQR de diferentes dimensiones sobre un mapa de la ciudad. Estos clústeres son generados por el algoritmo DBSCAN (cuyo funcionamiento fue explicado en la Sección 2.3.2.1), donde las agrupaciones son áreas con mayor densidad de elementos que otras áreas.

Para la creación y visualización del mapa se usa la misma directiva de Mapbox que usan los mapas de calor. Al igual que estos, los mapas de agrupaciones permiten visualizar clústeres generados sobre el conjunto total de servicios o para un servicio específico. Además, existe la opción de mostrar aquellos puntos que han quedado fuera de las agrupaciones (*outliers*), visualizados como marcadores individuales sobre el mapa. La herramienta también permite visualizar grupos de reportes en distintos niveles de granularidad. Estas granularidades dependen de diferentes valores de parámetros asociados con el algoritmo DBSCAN, como la distancia para calcular una vecindad y el número de puntos necesarios para que una región se considere densa, que dependerán de la cantidad de reclamos con los que se quieran computar clústeres. Estos valores están definidos en un archivo JSON en el servidor en donde se determinará una correspondencia entre la cantidad de reclamos y los parámetros de DBSCAN asociados.

Los clústeres en el mapa de la ciudad están codificados por colores para indicar si incluyen un número de puntos grande (rojo), medio (amarillo) o pequeño (verde). Al seleccionar un clúster se mostrará un polígono que determina el límite geográfico de esa agrupación. Además, la herramienta proporciona un mecanismo a través del cual un usuario puede visualizar subgrupos de la agrupación seleccionada. Al ir haciendo clic sobre diferentes subagrupaciones, se anidarán los polígonos que marcan límites geográficos. Para poder

realizar esto, se utiliza el algoritmo *hierarchical greedy clustering* (ver Sección 2.3.2.1) a fin de identificar los clústeres y subclústeres en diferentes niveles de zoom en un mapa. De este modo, el algoritmo DBSCAN se utiliza para identificar los grupos principales, y luego se combina con el método de agrupamiento jerárquico codicioso ofrecido por Mapbox para identificar de forma dinámica y eficiente subgrupos en diferentes niveles de granularidad.

Aparte del mapa de la ciudad, la página web tiene una sección dedicada a presentar información adicional dependiendo de la interacción con el mapa. En principio se indica el servicio y nivel de granularidad elegido, y el número de problemas y de reclamos, donde un mismo problema puede tener varios reclamos asociados. Al elegir un clúster, se muestra un gráfico de torta que indica las proporciones de reportes que pertenecen a diferentes categorías del servicio bajo análisis. Para generar este gráfico, se obtendrán y recorrerán los reportes de la agrupación para generar dinámicamente un archivo JSON con el formato especificado por la visualización elegida. En cambio, al seleccionar un reclamo individual (haciendo clic en el marcador que representa el reclamo), se puede ver información como la ubicación y los comentarios dejados por los usuarios respecto a ese reclamo.

4.4. Diseño

En esta sección se presentan las principales decisiones que se tomaron para el diseño del sistema, las cuales se explican en la Sección 4.4.1, indicando la arquitectura del sistema, y en la Sección 4.4.2 que presenta el modelo de datos.

4.4.1. Diseño de la aplicación

El diseño de la estructura se basa en el estilo cliente-servidor (BCK03). Fue elegida porque permite la centralización del control (accesos, recursos e integridad de datos) en el servidor; por su escalabilidad (tanto de aumentar el número de clientes y servidores dinámicamente y según se requiera) y por su facilidad de mantenimiento. Los componentes de la herramienta pueden observarse en el diagrama de componentes presente en la Figura 4.2. Aquellos presentes en el paquete *Frontend* se ejecutan en el lado del cliente, mientras que el resto se ejecuta en el servidor.

El diseño del sistema tiene tres paquetes principales: *Cliente*, *Servidor* y *Datos*. El *Cliente* provee los componentes que implementan todas las funcionalidades del sistema

que pueden accederse mediante el navegador. La interfaz de usuario está compuesta por tres componentes: aquel para observar la visualización de datos estadísticos sobre los servicios (*UI Datos estadísticos*), el que permite visualizar los mapas de calor (*UI Mapa de calor*) y el que permite visualizar los mapas de agrupaciones (*UI Mapa de agrupaciones*).

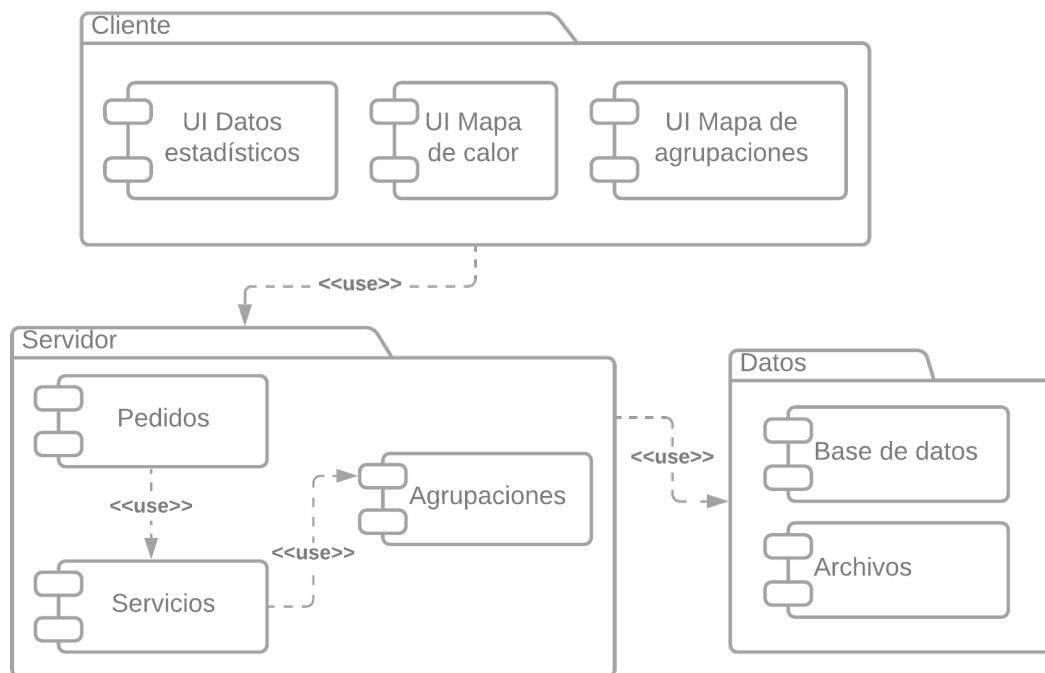


Figura 4.2: Diagrama de componentes.

Por su lado, el paquete *Servidor* implementa el componente que provee la funcionalidad principal y los servicios del sistema, encargado de tomar los pedidos del *Cliente* y devolver los datos esperados (componente *Pedidos*), y realizar las tareas de generar los archivos requeridos para los mapas de agrupaciones y los mapas de calor (componente *Servicios*). Estas funciones no se ejecutan a pedido del usuario, sino que lo hacen cada cierto período de tiempo y en *background*, a fin de actualizar los datos de los reportes que se reflejarán en las diferentes visualizaciones. Para determinar cada cuánto se ejecutarán estas rutinas del servidor, deberá tenerse en cuenta tanto la fluctuación en la generación y resolución de reclamos como la importancia de visualizar información más actualizada. El componente *Agrupaciones* es usado por el componente *Servicios* para utilizar el algoritmo DBSCAN de ELKI y así generar las agrupaciones de reportes para los mapas de agrupaciones.

El paquete *Datos* contiene los componentes donde se guardarán los datos y archivos

temporales. Todos los archivos generados por los componentes *Agrupaciones* y *Servicios* se guardarán en el componente *Archivos*, que luego serán accedidos por diferentes funciones del *Servidor*. Finalmente, el componente *Base de datos* es el responsable del manejo de la base de datos, incluidas las operaciones de guardar, obtener y borrar elementos. El diseño de la base de datos se indica en la siguiente sección.

4.4.2. Modelo de datos

El conjunto de datos que pueden utilizarse en la herramienta debe estar conformado por los reclamos de los ciudadanos. En particular, es una base de datos estructurada que debe tener los siguientes campos:

- *ID*: corresponde al identificador del PQR en el conjunto de datos.
- *ID_Servicio*: corresponde a un número o un texto que identifica un servicio específico. Por ejemplo, algunos servicios pueden ser: «Luminaria», «Arbolado Urbano», «Higiene Urbana», «Ingeniería vial», «Señalamiento vial», entre otras.
- *Categoría*: corresponde a la subcategoría dentro de un dado servicio.
- *Comentario*: es el comentario presentado por el ciudadano al realizar el reclamo.
- *Calle y Altura*: determinan la ubicación del reporte en la ciudad.
- *Latitud y Longitud*: las coordenadas geográficas en donde se encuentra la problemática reportada por el ciudadano.

Adicionalmente, cada entrada puede tener los siguientes campos:

- *Fecha y Hora*: corresponde a la fecha y hora en la cual se realizó el reclamo.
- *Canal*: corresponde al canal por el cual se efectuó el reclamo. Puede ser mediante la web, la aplicación o por llamada telefónica.
- *Género*: corresponde al género de quien reporta el problema. Puede ser masculino, femenino o desconocido, si no fue especificado.

Tabla 4.1: Ejemplo de una entrada en la base de datos provista por Citymis.

| ID | ID Servicio | Categoría | Comentario | Calle | Altura | Latitud | Longitud |
|--------|-------------|-----------|----------------------|-------|--------|----------------------|----------------------|
| 602159 | 1 | Apagada | Está siempre apagada | CUYO | 2051 | -34.495271 410000 | -58.526526 340000 |

La Tabla 4.1 presenta un ejemplo de una entrada en la base de datos provista por Mismatica². El *ID_Servicio* 1 corresponde al servicio de «Luminaria», siendo «Apagada» un problema o subcategoría que puede presentarse para ese servicio. Por otro lado, el reclamo está ubicado en «Cuyo 2051», siendo la latitud y longitud presentadas las asociadas a esa dirección dentro de la ciudad. Además, el ciudadano que realizó el reclamo agregó un comentario diciendo «Está siempre apagada» para agregar información respecto al problema. Finalmente, este nuevo reclamo tomó el identificador 602159.

En el desarrollo de esta herramienta se utilizaron datos de Citymis Community³ y de Gestión Colaborativa⁴. Ambas plataformas se presentarán en la Sección 4.6 y Sección 4.7 respectivamente. El primer conjunto de datos fue provisto por la empresa Mismatica y corresponde a reportes generados por usuarios de San Isidro por medio de la plataforma Citymis Community en el año 2016. El mismo consiste de 68.368 reportes realizados por medio de la aplicación móvil, el sitio web o por medio de un número telefónico. El segundo se obtuvo de la página de datos abiertos de Buenos Aires y corresponden a reportes generados por medio de la plataforma Gestión Colaborativa de la Ciudad de Buenos Aires. Los datos agrupan 40.656 reportes realizados por ciudadanos mediante la página web, la aplicación BA147⁵ o el teléfono gratuito para reportar problemáticas durante el año 2017⁶.

4.5. Implementación y testeo

Para el desarrollo de las funciones provistas por el *frontend* se hizo uso de HTML, CSS y el lenguaje de programación JavaScript⁷, utilizándose también el framework Bootstrap⁸. Las distintas facilidades del cliente requerirán de datos provistos por el lado servidor, desarrollado en el lenguaje de programación PHP⁹, que se encuentra encargado de la manipulación de la información asociada a los reportes ciudadanos, usándose MySQL¹⁰

² Mismatica - <https://mismatica.com/es/index.php>

³ Citymis Community - <https://mismatica.com/clientes/propuestas/tecnic/master/>

⁴ Gestión Colaborativa - <https://gestioncolaborativa.buenosaires.gob.ar/prestaciones>

⁵ BA147 - <https://www.buenosaires.gob.ar/aplicacionesmoviles/ba-147>

⁶ Dataset 2017 - <https://data.buenosaires.gob.ar/dataset/>

[sistema-unico-atencion-ciudadana/archivo/juqdkmgo-1958-resource-xlsx](#)

⁷ JavaScript - <https://www.javascript.com/>

⁸ Bootstrap - <http://getbootstrap.com/>

⁹ PHP - <http://php.net/>

¹⁰ MySQL - <https://www.mysql.com/>

para la implementación de la base de datos.

Para visualizar los mapas de la región geográfica y representar los reportes se utilizó la herramienta Mapbox¹¹. Por otro lado, se empleó D3¹² y amCharts¹³ para la creación de gráficos que acompañan datos estadísticos sobre los reportes, y servicios y categorías asociadas a ellos. También se hizo uso de las competencias del framework ELKI¹⁴ para la ejecución de algoritmos de minería de datos sobre las características de los reportes, en particular, sobre sus ubicaciones geográficas. Los pares *[latitud, longitud]* asociados a cada reclamo se utilizaron para la ejecución en diferentes instancias del algoritmo DBSCAN, explicado en la Sección 2.3.2.1.

Mapbox es una plataforma que se especializa en la cartografía digital, otorgando no sólo la visualización de mapas y la posibilidad de diseñarlo de manera arbitraria, sino también gran variedad de facilidades para la visualización de diferentes objetos, de interés para el desarrollador, sobre los mapas que provee. En particular, se hablarán de las facilidades que se utilizaron en el desarrollo de la página web, siendo estas principalmente:

- la visualización del mapa de la ciudad, que puede hacerse mediante una instrucción dada por la API de Mapbox;
- *heatmap*¹⁵ para la realización de un mapa de calor para los reportes; y
- *markercluster*¹⁶, para la visualización de las agrupaciones de reportes generados por el algoritmo DBSCAN comentado anteriormente.

Para representar los objetos sobre el mapa creado, objetos como el marcador de una dada ubicación en la ciudad o diferentes figuras geométricas, las funciones mencionadas requieren antes obtener información geográfica sobre ellos, y, opcionalmente, datos secundarios, como puede ser un título. Para otorgarle esta información, se optó por hacer uso de archivos GeoJSON¹⁷.

Los testeos se realizaron a nivel de pruebas unitarias por la tesista, mientras que las directoras contribuyeron a verificar pruebas de integración y de uso. Los testeos se

¹¹ Mapbox - <https://www.mapbox.com/>

¹² D3 - <https://d3js.org/>

¹³ amCharts - <https://www.amcharts.com/>

¹⁴ ELKI - <http://elki.dbs.ifi.lmu.de/>

¹⁵ heatmap - <https://www.mapbox.com/mapbox-gl-js/example/heatmap/>

¹⁶ markercluster - <https://www.mapbox.com/mapbox.js/example/v1.0.0/>

[markercluster-with-mapbox-data/](#)

¹⁷ GeoJSON - <http://geojson.org/>

complementaron con demostraciones del sistema realizados ante partes interesadas en el desarrollo de la herramienta.

4.6. Caso de estudio: Citymis

El primer caso de estudio utiliza datos provistos por Mismatica², recopilados mediante su aplicación Citymis Community³ durante el año 2016. Citymis es un ecosistema de aplicaciones web y móviles para la gestión integral de los servicios municipales, diseñado tanto para ciudadanos como para agentes públicos municipales, que permite a los usuarios informar problemas directamente desde sus computadoras o teléfonos inteligentes.

Para describir el flujo de mensajes y/o tareas entre el ciudadano y el funcionario público hasta la resolución del problema, se comenzará desde el punto en el que se realiza el reclamo de una problemática. Al momento de realizar un reporte, el usuario puede distinguirlo especificando una categoría del problema. La lista de categorías definidas por Citymis en la base de datos utilizada es la siguiente: «Luminaria», «Arbolado urbano», «Ingeniería vial», «Higiene urbana», «Señalamiento vial», «Bromatología», «Parques y plazas» y «Semáforos». Es destacable mencionar que Citymis provee una gran variedad de servicios que exceden a los nombrados. A su vez, cada una de estas categorías tiene una serie de subcategorías para una mejor especificación de la problemática observada. Por ejemplo, dentro de «Luminaria» se incluyen categorías relacionados a luminarias apagadas e intermitentes, pedido de una nueva luminaria o pedido de reparación, entre otras. El ciudadano también deberá ingresar la ubicación, identificada con la calle y el número, donde se encuentra el problema o, alternativamente, especificarla por medio de un mapa. La existencia del mapa no solo provee la facilidad de especificar la ubicación geográfica exacta, sino también de encontrar y examinar otros problemas reportados en la misma área. Además, el usuario puede escribir un comentario y proporcionar una foto para una mejor descripción del problema. La aplicación también requiere de información de contacto a fin de poder notificar al usuario sobre cualquier posible cambio en el estado del reclamo.

Desde una perspectiva operacional, la municipalidad se divide en subregiones geográficas llamadas «áreas de asignación». Esta división permite asignar los reportes a empleados específicos (inspector, equipo de mantenimiento, etc.) para resolver o verificar los problemas reportados. Cada reclamo es asociado en primera instancia a una «unidad de gestión»,

que es básicamente una ubicación geográfica (latitud, longitud, calle y altura) y se le asigna el estado de «pendiente de moderación». En este punto, el reporte puede ser aceptado, en cuyo caso se le cambia el estado a «recibido», o puede ser descartado, recibiendo el estado de «rechazado». Una vez que el reclamo es aceptado, el siguiente paso es resolver el problema a fin de obtener el estado de «cumplido» y terminar el proceso. Cada vez que cambia el estado de un reclamo, el usuario que realizó el reporte recibe una notificación por correo electrónico y es invitado a realizar nuevos comentarios para brindar información adicional que pueda ser de ayuda para resolver el problema, o simplemente para dar su opinión respecto al servicio.

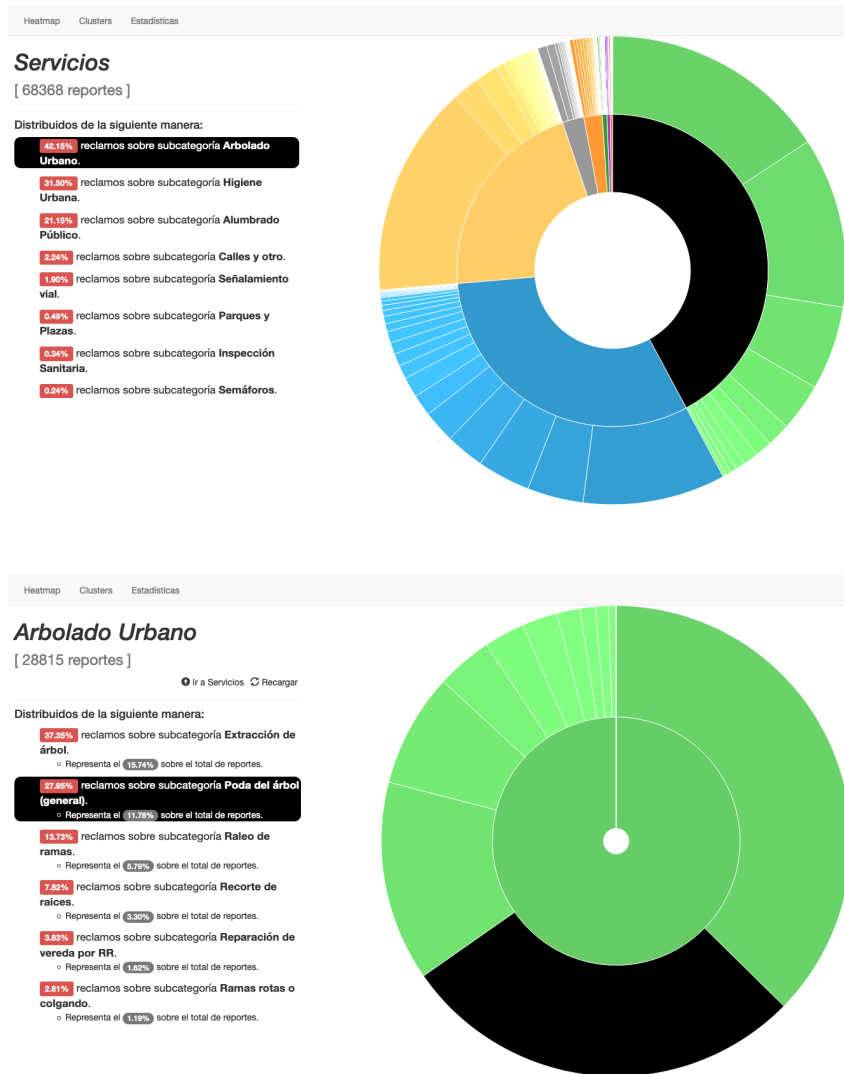
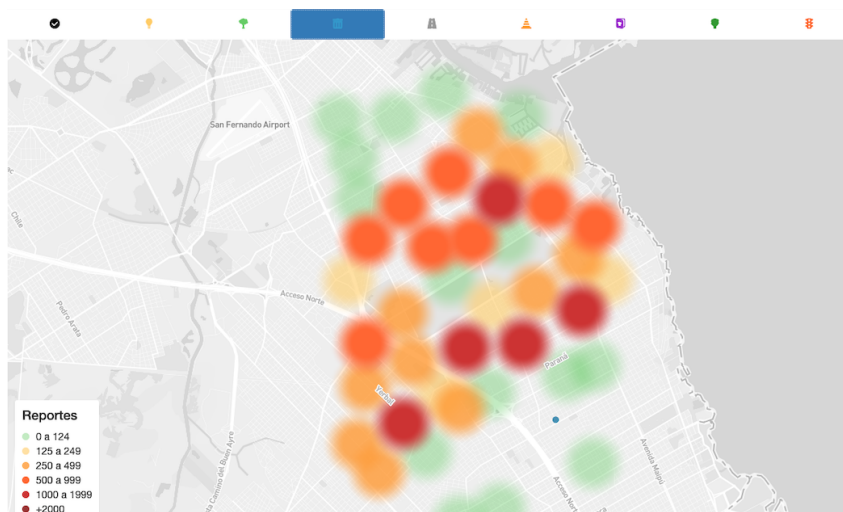


Figura 4.3: Funcionalidad de navegación de datos estadísticos.

Se ilustrará el uso de la herramienta con los datos provistos por Citymis, mostrando el uso de cada funcionalidad de la herramienta con un servicio de los más afectados de la ciudad. Para empezar a analizar las comunicaciones ciudadanas es ideal la funcionalidad de datos estadísticos, dado que permite explorar las proporciones de servicios y categorías más afectadas, dando una rápida idea inicial de los servicios que requieren más atención. La Figura 4.3 presenta un gráfico de proyección solar con las proporciones de reclamos en los diferentes servicios, siendo los más afectados «Arbolado urbano» (en verde), «Higiene urbana» (en azul) y «Alumbrado público» (en amarillo). La cantidad de reclamos en el resto de los servicios es considerablemente menor. Se puede navegar en el gráfico para observar la cantidad de reportes en diferentes categorías dentro de un servicio, como puede verse en la segunda subfigura. En este caso, dentro del servicio de «Arbolado urbano», las categorías más reclamadas están asociadas a la extracción o la poda de árboles, seguido con recortes de ramas o raíces.

La Figura 4.4 muestra el mapa de calor para los reportes registrados del servicio «Higiene urbana». En la primera imagen puede verse en totalidad la ciudad de San Isidro y el mapa de calor para el servicio mencionado. En particular, pueden observarse algunas zonas más afectadas por problemáticas (en colores bordó y anaranjado). La segunda imagen presenta el mapa luego de realizar zoom sobre la zona norte de la ciudad. Allí se pueden observar concentraciones de reportes de diferentes tamaños, haciendo más fácil la localización de zonas más pequeñas con mayor concentración de reportes. Se puede realizar más zoom, como muestra la tercera figura, que permite observar puntos azules



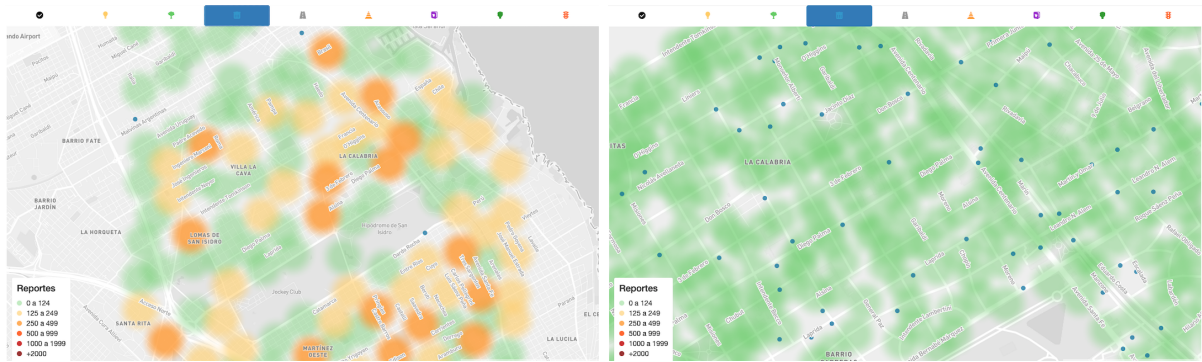
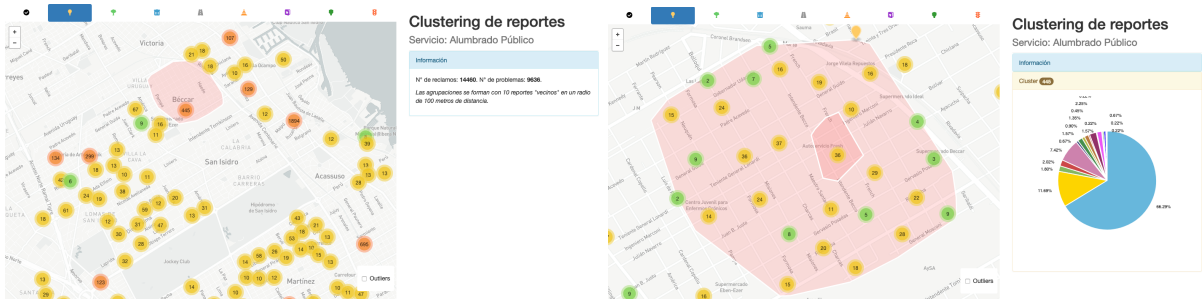


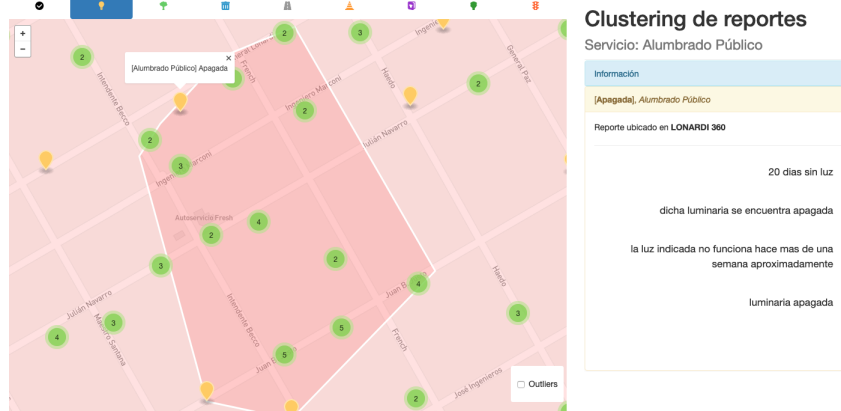
Figura 4.4: Mapa de calor para el servicio «Higiene Urbana» en la ciudad de San Isidro a diferentes niveles de zoom.

como los reclamos específicos realizados por los ciudadanos.

El mapa de calor ayuda a una fácil y rápida identificación de la zonas más afectadas, pero su análisis puede profundizarse con el mapa de agrupaciones. La Figura 4.5 muestra el mapa de clústeres para el servicio «Alumbrado público». En la subfigura 4.5a pueden observarse las diferentes agrupaciones en el mapa, la información acerca de reclamos y problemas, y cómo se ve el polígono asociado al clúster al pasar el mouse por encima. El resultado de explorar esa agrupación haciendo clic sobre ella puede apreciarse en la subfigura 4.5b. Aquí se puede observar en más detalle la zona geográfica que abarca la agrupación, los subclústeres que la conforman, e información adicional acerca de la agrupación en el gráfico de torta. Este gráfico indica las proporciones de reclamos en las diferentes categorías del servicio de «Alumbrado público». En este caso, más de la mitad de los reclamos dentro de esa zona geográfica corresponden a luminarias apagadas. En esta imagen también se observa que es posible tener una serie de polígonos anidados al ir seleccionando subclústeres dentro de una agrupación. Al explorar lo suficiente las



(a) Visualización del polígono de un clúster al pasar el mouse por encima. (b) Polígonos anidados de agrupaciones e información adicional sobre el clúster.



(c) Información adicional al seleccionar un problema.

Figura 4.5: Mapa de agrupaciones para el servicio «Alumbrado Público» en la ciudad de San Isidro.

agrupaciones aparecerán los marcadores que indican los problemas en la zona. Al realizar clic sobre el marcador se podrá ver información asociada al problema, como la calle y número donde se encuentra ubicado y los comentarios de los vecinos. En la subfigura 4.5c se observa un ejemplo de reclamos y comentarios respecto a una luminaria apagada.

4.7. Caso de estudio: Gestión colaborativa

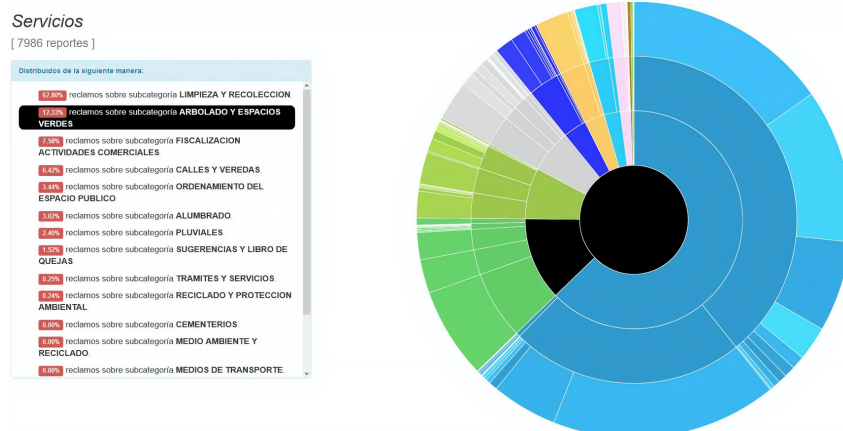
El segundo caso de estudio utiliza datos recuperados de la página de datos abiertos de Buenos Aires correspondientes a los reportes realizados en el año 2017 mediante la aplicación Gestión Colaborativa⁴. «Gestión colaborativa» es una plataforma digital que está destinada a la atención eficiente de PQR realizados por los ciudadanos respecto a los servicios que brinda la Ciudad Autónoma de Buenos Aires.

Describiremos ahora los pasos que debe realizar un usuario si desea reportar una problemática por medio de la web o la aplicación BA147, y las facilidades que proveen dichos canales. Para realizar un reclamo, el ciudadano debe identificarse en la aplicación web con su cuenta miBA¹⁸. Esa cuenta es el perfil del ciudadano en la Ciudad, siendo la plataforma el lugar desde donde puede gestionar trámites y servicios. El registro en la plataforma puede hacerse por medio de su vinculación con Facebook o Google, o por medio de usuario y contraseña. Una vez identificado, el usuario tiene la opción de efectuar un

¹⁸ miBA - <https://miba.buenosaires.gob.ar>

nuevo reclamo, eligiendo el tipo de servicio sobre el que quiere reclamar. El tipo de servicio puede ser uno de los siguientes: «Alumbrado», «Arbolado y espacios verdes», «Calles y veredas», «Control edilicio, obras y catastro», «Fiscalización de actividades comerciales», «Ordenamiento de espacio público», «Limpieza y recolección», «Pluviales», «Reciclado y protección ambiental», «Salud y servicios sociales», «Seguridad», «Trámites y servicios» y «Tránsito y transporte». Para cada una de estos servicios, hay categorías para especificar la problemática observada. Por ejemplo, las categorías dentro del servicio «Alumbrado» están relacionadas con reparaciones de luminarias o pedido de mayor iluminación en una calle, entre otras. Existe un tercer nivel de especificación, los rubros dentro de cada categoría, que también pueden especificarse al momento de realizar un reclamo.

Además de seleccionar servicio y categoría, el ciudadano deberá ingresar la calle y el número donde se encuentra ubicado el problema, o alternativamente especificar la ubicación usando el mapa que ofrece la aplicación. El usuario también puede agregar un comentario y una foto para mejorar la descripción del problema. Una vez realizados estos pasos, debe confirmar el reporte. La aplicación mantiene informado al ciudadano acerca de cualquier cambio en el estado de la solicitud. El mapa presente en la aplicación permite además observar todos los reclamos activos realizados por otros usuarios. Si ya existe el reclamo que se intenta hacer (mismo servicio, categoría y ubicación), el usuario tiene la oportunidad de reforzar el reclamo ya existente, aumentando su prioridad. Esto puede hacerse con cualquier otro reporte que figure activo en el mapa. Adicionalmente, la aplicación permite actualizar el estado de un reclamo con fotos o comentarios. Cuando la solicitud es finalmente atendida y cumplida, el usuario recibe un correo electrónico con los detalles del trabajo realizado y una encuesta de satisfacción respecto al servicio prestado.



(a) Datos estadísticos de todos los servicios.

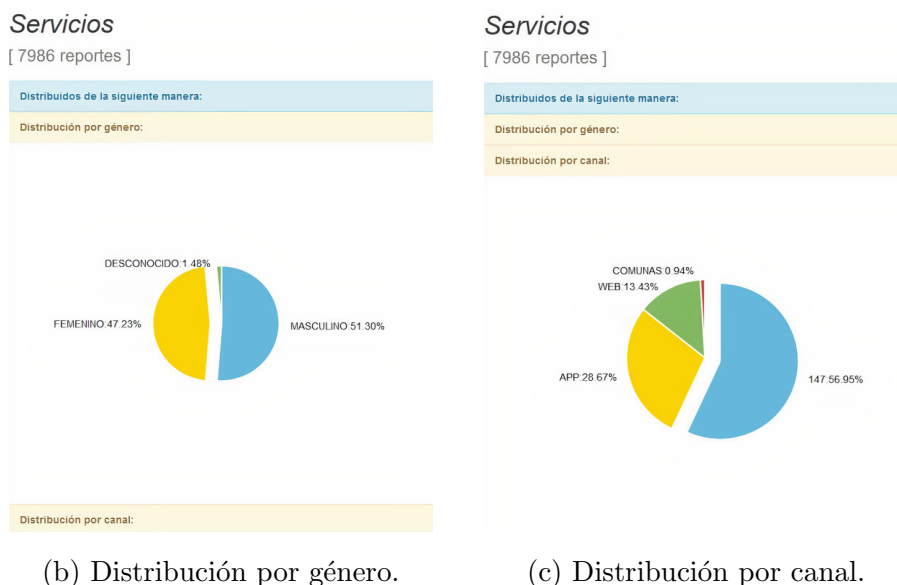


Figura 4.6: Funcionalidad de navegación de datos estadísticos.

Para explorar las funcionalidades de la herramienta desarrollada, se utilizarán los datos recopilados mediante la aplicación Gestión Colaborativa, en particular aquellos reportes realizados durante el año 2017. Para estos datos se utiliza una segunda versión de la herramienta, que incorpora dos nuevas secciones en la funcionalidad de datos estadísticos, haciendo un mejor uso de este conjunto de datos. La Figura 4.6a muestra el gráfico de proyección solar para todos los servicios de la Ciudad de Buenos Aires. A diferencia del municipio de San Isidro, en CABA el servicio más reclamado es «Limpieza y recolección», conformando 62,80 % de los reportes ciudadanos. Otros servicios afectados son «Arbolado

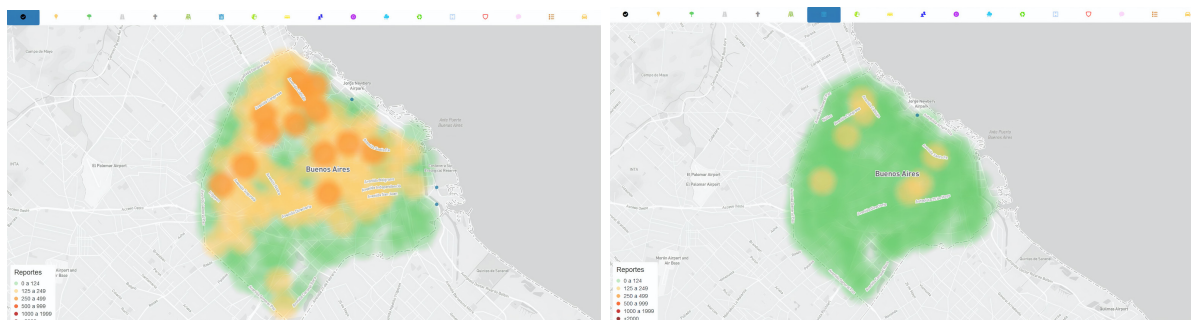
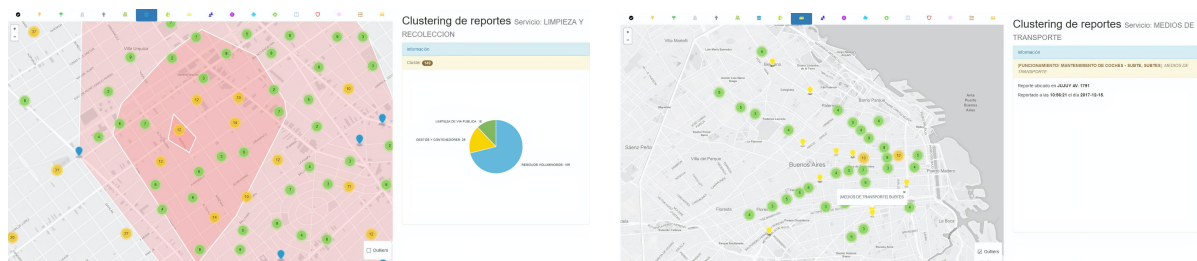


Figura 4.7: Mapas de calor para la Ciudad de Buenos Aires.

y espacios verdes» y «Fiscalización de actividades comerciales». Las nuevas secciones de la herramienta pueden verse en las siguientes dos figuras. La Figura 4.6b sugiere que el género no tiene incidencia en los reportes generados. Por otro lado, la Figura 4.6c indica que los medios más utilizados para realizar reclamos son principalmente llamadas telefónicas al número gratuito 147 y la aplicación móvil.

La Figura 4.7a muestra el mapa de calor para la totalidad de los servicios ofrecidos por la Ciudad de Buenos Aires. Pueden observarse algunos puntos más afectados en la ciudad, sobre todo en la parte centro-norte. Por otro lado, la Figura 4.7b presenta el mapa de calor para el servicio «Limpieza y recolección», siendo el servicio con más reclamos en CABA. Sin embargo, los reclamos parecen bien distribuidos a lo largo de toda la ciudad, habiendo solo cinco puntos de mayor concentración. Sería de utilidad para funcionarios públicos explorar en más detalle estos puntos de mayor concentración de reportes.

En la Figura 4.8a puede verse una serie de polígonos anidados en una parte afectada de la ciudad respecto al servicio «Limpieza y recolección». Para esta agrupación en particular, la categoría más reclamada es «Residuos voluminosos», como indica el gráfico de torta en el lado derecho de la imagen. Por otro lado, la Figura 4.8b muestra reclamos asociados al servicio «Medios de transporte». Como puede apreciarse en el mapa de la ciudad, este servicio no está tan afectado como otros, como indican las agrupaciones de pocos reclamos (de colores verdes en su mayoría). En la figura también se observa un reporte seleccionado y su información asociada, como ubicación y fecha y hora de generación del reclamo. Este conjunto de datos no tiene comentarios asociados a los reportes.



(a) Polígonos anidados para el servicio «Limpieza y recolección».

(b) Selección de un reporte del servicio «Medios de transporte».

Figura 4.8: Mapa de agrupaciones para la Ciudad de Buenos Aires.

4.8. Discusión

En este capítulo se presentó un prototipo que ayuda a identificar los problemas predominantes en un municipio. La herramienta apoya el análisis de PQR ciudadanas mediante la aplicación de métodos basados en técnicas de agrupación y visualización. Estos métodos incluyen el uso de mapas de calor para representar la intensidad de los informes en puntos geográficos, la aplicación de agrupaciones jerárquicas y basadas en la densidad para visualizar clústeres de informes en la ciudad, y el uso de un gráfico de proyección solar para producir representaciones navegables de estadísticas obtenidas de los datos recopilados.

El mapa de calor ayuda a distinguir distintas concentraciones de reportes a lo largo de diferentes zonas geográficas, haciendo más simple el trabajo de reconocer las áreas más afectadas. Así mismo, el poder enfocarse en ciertos tipos de servicios puede ayudar a identificar problemas y patrones predominantes relacionados con esos servicios. La funcionalidad para observar agrupaciones de reportes y sus límites en el mapa ofrece una valiosa ayuda para definir estrategias que mejoren los niveles de coordinación de servicios. En particular, permite diseñar una mejor logística para atender las demandas de los ciudadanos considerando el tipo de servicio y su cercanía geográfica. El uso del gráfico de proyección solar resulta una manera intuitiva para presentar estadísticas en una estructura jerárquica visual que es fácil de navegar. En consecuencia, ofrece una herramienta adecuada para tener una visión clara de las dimensiones de los problemas de un municipio en función de sus correspondientes categorías y subcategorías de servicios.

A través de estas funcionalidades, la herramienta ofrece una valiosa ayuda para asistir al reconocimiento de las áreas más afectadas, a la planificación logística a fin de atender las demandas ciudadanas de manera eficiente, al diseño de estrategias para mejorar oportuna y proactivamente la infraestructura de servicios en una comunidad, a la identificación cuantitativa de las relaciones causa-efecto resultantes de las políticas públicas del municipio mediante la evaluación de acciones concretas en áreas geográficas delimitadas, y a la elaboración de encuestas para mantener informados a la ciudadanía mediante datos visuales y estadísticas. El uso de la herramienta puede conducir a una organización interna más eficiente al reducir los costos de mantenimiento. También puede resultar en un mejor uso de los recursos públicos al conducir a decisiones más inteligentes y mejores inversiones con efectos cuantitativos medibles. Finalmente, puede ayudar a mejorar la comunicación política y, por lo tanto, a desarrollar el entendimiento entre los ciudadanos y los funcionarios del gobierno.

La herramienta presenta una serie de limitaciones y abre la posibilidad a varias actualizaciones como parte de trabajo a futuro. Respecto a las limitaciones, actualmente la herramienta utiliza algoritmos que requieren ejecutarse en *background* como una rutina separada y previa a los pedidos del usuario. Sería adecuado actualizar estos algoritmos para que se ejecuten a demanda, de manera de presentar al usuario los datos de la manera más actualizada posible. Entre las mejoras, se planea incluir una dimensión temporal de los algoritmos de agrupación para permitir el análisis de informes en una ventana de tiempo específica. Esto haría posible visualizar PQR generados, por ejemplo, durante un mes específico o una semana en particular. También se considera ampliar el análisis que ofrece la herramienta considerando fuentes adicionales de datos como opiniones ciudadanas reflejadas en redes sociales o comentarios de lectores de periódicos digitales. La incorporación de estos nuevos datos abre nuevos desafíos de investigación, como el desarrollo de métodos personalizados para búsqueda basada en temas, reconocimiento de entidades nombradas, análisis de sentimientos, etc.

Capítulo 5

Herramientas informáticas para el análisis de interacción entre gobierno y ciudadanos

En este capítulo se presenta una plataforma que aplica técnicas de análisis de sentimiento y diferentes visualizaciones para analizar las interacciones que tienen lugar en Twitter entre el gobierno y los ciudadanos. La motivación se presenta en la Sección 5.1. En la Sección 5.2 se muestra la metodología de desarrollo, y en la Sección 5.3 los requerimientos y funcionalidades de la herramienta. El diseño, la implementación y el testeo se indica en las secciones 4.4 y 4.5. La validación de la herramienta se realiza mediante dos casos de estudio, que se presentan en las secciones 5.6 y 5.7. Finalmente, en la Sección 5.8 se discute las contribuciones y las limitaciones de la plataforma desarrollada.

5.1. Motivación

Las redes sociales, particularmente Twitter, se han convertido en una herramienta ampliamente utilizada por los gobiernos para compartir noticias, brindar información sobre servicios públicos, hacer anuncios de políticas y, en general, interactuar con los ciudadanos y evaluar sus puntos de vista sobre las políticas y el desempeño del gobierno. A través de Twitter, los ciudadanos pueden expresar sus opiniones sobre las decisiones y anuncios del gobierno, ya sea dando me gusta a un tweet, compartiendo o respondiendo a una publicación, etc. A medida que dichas interacciones se vuelven más ricas en contenido y formas

de interactuar, se convierten en una fuente importante de información para la toma de decisiones del gobierno. Las técnicas de minería de texto y visualización de datos ofrecen muchas oportunidades para hacer que dicha información sea comprensible e interpretable para los responsables de la formulación de políticas, los expertos en el dominio y otros usuarios no técnicos. A pesar de esto, existe una notable falta de herramientas de software listas para usar a fin de procesar la información publicada y las interacciones que tienen lugar en redes sociales, con el objetivo último de ayudar en la toma de decisiones.

Este estudio está impulsado por la pregunta de investigación definida previamente en la Sección 1.2: **¿Qué componentes y técnicas pueden ser parte de una infraestructura de software que haga uso de información volcada en plataformas digitales para facilitar la toma de decisiones por parte de funcionarios públicos?** Se busca el diseño e implementación de una plataforma que facilite el análisis de las interacciones gobierno-ciudadano que tienen lugar en redes sociales, utilizando las técnicas de minería de datos y de visualización adecuadas al dominio. En particular, para abordar cuestiones como (1) el nivel de actividad por parte de los gobiernos en diferentes países y sectores, (2) los temas que se abordan, (3) los recursos compartidos como parte de la difusión de mensajes, (4) la intensidad de la respuesta ciudadana a las publicaciones del gobierno, (5) el sentimiento predominante expresado por los ciudadanos en esas respuestas, y (6) las posibles combinaciones de lo anterior.

Se presentará a continuación una herramienta de software para ayudar a los usuarios expertos y no expertos a analizar las publicaciones gubernamentales y las opiniones de los ciudadanos en redes sociales, en particular en Twitter. La herramienta implementa una serie de técnicas de análisis de datos y de visualización que ayudan a analizar cómo los gobiernos utilizan Twitter para comunicarse con los ciudadanos y cómo los ciudadanos responden a dicha comunicación. Las funcionalidades y técnicas utilizadas se describirán en secciones siguientes.

5.2. Metodología de desarrollo

La implementación de la herramienta siguió una metodología de desarrollo evolutiva, cuyas características principales y la justificación de su uso se explicó en la Sección 4.2. Al igual que la herramienta anterior, los requerimientos se plantearon para responder a la pregunta de investigación, de modo que se definieron diversas funcionalidades, que fueron

luego implementadas y mejoradas en los diferentes ciclos del desarrollo. En las siguientes secciones se presentan las características más importantes del desarrollo.

5.3. Requerimientos

Las principales funcionalidades de la herramienta se pueden observar en el diagrama de caso de uso presentado en la Figura 5.1. Los usuarios de la herramienta puede ser un oficial de comunicaciones, un oficial de políticas públicas o un ciudadano. Las funcionalidades del sistema son: *Analizar uso de recursos multimedia*, *Analizar interacciones ciudadanas*, *Analizar sentimientos*, *Identificar temas destacados* y *Analizar actividad*. Todas estas funcionalidades también implican utilizar la funcionalidad *Obtener información de cuenta*, para seguir o dejar de seguir cuentas de interés. Algunas utilizan la funcionalidad *Recuperar noticia* para obtener la información y comentarios de una noticia. El detalle de todas las funcionalidades será abordado en las siguientes secciones.

El ciudadano puede estar interesado en el uso de la herramienta para observar la actividad de una cuenta de gobierno (por ejemplo, de algún sector público) y determinar cuáles son los temas tratados. La información que provee la herramienta respecto a los temas tratados y las respuestas obtenidas de los ciudadanos (tanto en términos de interacciones como de sentimientos identificados) pueden ser utilizadas por un oficial de políticas públicas, que lo incluirá en su análisis y posterior toma de decisiones. Por su lado, un oficial de comunicaciones puede determinar nuevas estrategias de comunicación teniendo en cuenta la respuesta que obtiene de los ciudadanos y el actual uso de recursos multimedia por parte de la cuenta analizada.

5.3.1. Obtener información de cuenta

La herramienta permite al usuario «seguir» cuentas de diferentes redes sociales, en particular de Twitter. Cuando se realiza esta solicitud, el servidor chequea si tiene información de la cuenta en la base de datos. De ser así, se recuperan todas las publicaciones realizadas a partir de la fecha de la última publicación de la cuenta que se tenga en la base de datos; caso contrario, se recuperan las últimas cien publicaciones realizadas por la cuenta. El número de publicaciones recuperadas puede variar según las limitaciones

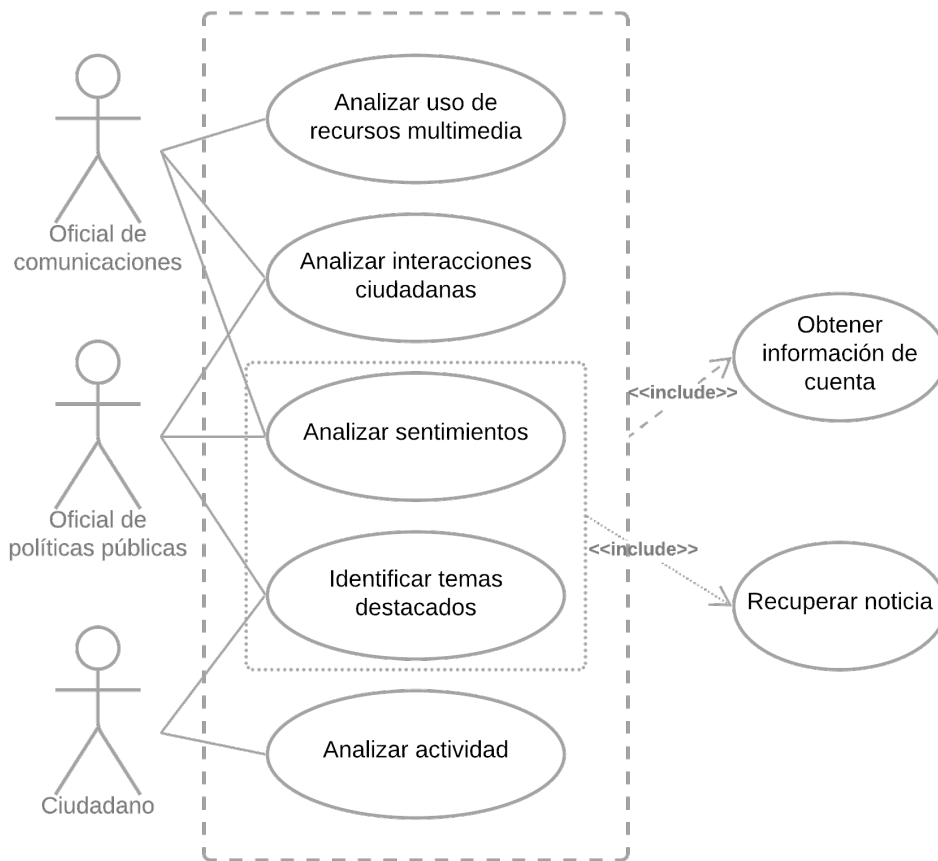


Figura 5.1: Diagrama de caso de uso.

impuestas por la API. Además, a partir de este momento, la herramienta estará controlando nuevas publicaciones realizadas por esa cuenta en tiempo real. Así mismo, el usuario puede dejar de seguir una cuenta si su contenido ya no es de interés; esto se registrará así en el servidor y se dejará de controlar su actividad.

5.3.2. Recuperar noticia

Una funcionalidad suplementaria de la herramienta permite al usuario recuperar información de una noticia, incluidos sus comentarios, en particular de los periódicos digitales Clarín, Página12, LaNueva y CBC Canada. Cuando se realiza esta solicitud, el servidor verifica si ya existe la noticia en la base de datos. Si es así, recupera todos los comentarios realizados en la noticia desde el último que tiene guardado; caso contrario, obtiene la información de la noticia y todos los comentarios actuales. Una vez obtenida esta información, la guardará en la base de datos y estará disponible para ciertas visualizaciones.

5.3.3. Analizar actividad

Un enfoque simple para analizar la presencia del gobierno en redes sociales es cuantificar los niveles de actividad en Twitter y otras redes. La visualización de los niveles de actividad ofrece información sobre las posibles desviaciones de los niveles normales de actividad. Este análisis combinado con el contexto social, político y cultural, ayuda a resaltar los factores relevantes para producir niveles altos de actividad e interacción, identificando también eventos de especial interés para el gobierno y la sociedad.

Para el análisis de actividad, la herramienta, en su primera iteración, proveía una visualización de *event drop*, tal como puede verse en la Figura 5.3. En su última versión, provee una visualización de gráfico de líneas, como los de la Figura 5.12. Para poder ver esta visualización, el usuario debe seleccionar una cuenta y luego la visualización deseada (*Time Series Chart*). Esto generará un pedido al servidor, que recuperará todos los tweets de la cuenta seleccionada. Con esos tweets se determina el número de tweets y los hashtags más utilizados por día. Así mismo, se devolverá información cuantitativa acerca de las interacciones (*retweets*, favoritos y comentarios) y de las emociones identificadas en los comentarios ciudadanos, conformando cada una de estas otras series de tiempo que podrán visualizarse junto con el número de publicaciones diarias.

5.3.4. Identificar temas destacados

Las nubes de palabras son una forma eficaz de resaltar los temas más destacados en un contexto determinado y son una herramienta útil para examinar una gran cantidad de tweets o de comentarios de una noticia. Esta es la visualización que provee la herramienta para la rápida identificación de los temas tratados. El usuario, previa selección de la cuenta de interés o de una noticia, puede solicitar la visualización *Word Cloud*. Esto generará un pedido al servidor, que obtendrá todos los tweets/comentarios de la cuenta/noticia elegida, y procesará su contenido para devolverlo. Este procesamiento implica identificar y contabilizar las palabras, eliminar las *stopwords*¹ y los signos de puntuación, y devolver las palabras con sus frecuencias y emociones asociadas.

¹Las *stopwords* son las palabras sin significado como artículos, preposiciones, pronombres, etc.

5.3.5. Analizar uso de recursos multimedia

El gobierno puede comunicarse con los ciudadanos a través de Twitter compartiendo diferentes tipos de recursos, como enlaces, imágenes y vídeos. El análisis del volumen de recursos de diferentes tipos compartidos por el gobierno en diferentes sectores revela los enfoques preferidos para hacer anuncios e involucrar a los ciudadanos. La herramienta ofrece como visualización los diagramas de Euler coloreados, donde la intensidad del color y el tamaño representan el grado de ocurrencia: cuanto mayor es la intensidad y el tamaño, más veces se comparte ese tipo de recurso.

El usuario puede solicitar esta visualización (*Euler Diagram*) previa selección de la cuenta de interés. Este pedido al servidor implica la recuperación de los tweets de la cuenta y el período seleccionado, y la identificación y cuantificación del uso de los recursos multimedia. También la contabilización de la cantidad de veces que se comparte más de un recurso multimedia en simultáneo.

5.3.6. Analizar interacciones ciudadanas

La medición del número de comentarios publicados por los ciudadanos en respuesta a las publicaciones del gobierno cuantifica el grado de participación ciudadana. Además, también se pueden medir y analizar otras formas de respuesta ciudadana, como marcar un tweet como favorito o compartirlo (*retweet*).

La herramienta provee dos visualizaciones que permiten identificar los niveles de interacción, el gráfico de serie de tiempo de la Subsección 5.3.3 y el diagrama de Hasse extendido. El usuario puede elegir esta última visualización luego de seleccionar la cuenta de interés. Este pedido al servidor implica: recuperar los tweets de la cuenta y período seleccionado, e identificar y cuantificar las posibles combinaciones de interacciones recibidas. Esto significa determinar el número de tweets que recibieron al menos una interacción, el número de tweets que recibieron cada tipo de interacción (comentarios, favoritos y *retweets*), y el número de tweets que recibieron más de un tipo de interacción simultáneamente (por ejemplo, cuántos tweets recibieron comentarios y fueron compartidos).

La herramienta también provee un diagrama de Sankey. Esta visualización permite observar la relación entre el uso de los recursos multimedia y la respuesta obtenida de los ciudadanos. Como el resto de los pedidos al servidor, el usuario puede elegir esta visualización (*Sankey Diagram*), previa selección de la cuenta de interés. El servidor recupera

los tweets de la cuenta seleccionada, y determina: (1) para la totalidad de tweets, los porcentajes de tweets que hacen uso de cada recurso multimedia y el porcentaje de tweets que solo utilizan texto, (2) para cada uno de los recursos multimedia x , el porcentaje de tweets que compartieron x y recibieron z interacción (z correspondiente a cada una de las interacciones posibles), y (3) para los tweets que solo utilizaron texto, el porcentaje de tweets que recibieron z interacción.

5.3.7. Analizar sentimientos

Poder reconocer las emociones de los ciudadanos ante diferentes eventos es de gran utilidad para tener una idea de la opinión pública general sobre ciertos temas. El análisis de sentimiento provisto por la herramienta se llevó a cabo sobre los comentarios de los ciudadanos y está basado en la rueda de emociones de Plutchik (Plu01). Según la teoría de Plutchik, hay ocho emociones principales: alegría, confianza, miedo, sorpresa, tristeza, disgusto, enojo y anticipación.

Una técnica sencilla para llevar a cabo el análisis de sentimientos en un conjunto de datos consiste en basarse en un léxico que capture las asociaciones palabra-emoción y contar el número de ocurrencias en el corpus asociadas con cada emoción. Esta técnica permite estimar cuáles son las emociones predominantes en el conjunto de datos analizado. En la herramienta, este análisis se llevó a cabo utilizando la versión en español del *NRC Affect Intensity Lexicon* (MT10), realizada con el traductor de Google². A modo de ilustración, la Tabla 5.1 contiene ejemplos de las palabras en español asociadas con las ocho emociones primarias.

La herramienta provee tres visualizaciones para analizar las emociones. La primera fue presentada en la Subsección 5.3.3, un gráfico de líneas con la posibilidad de observar la evolución de las emociones a lo largo del tiempo, superpuesto a la actividad de la cuenta analizada. Las otras dos visualizaciones se indicarán a continuación.

El usuario puede seleccionar una visualización de gráfico de radar, previa selección de la cuenta o noticia de interés, haciendo clic sobre *Radar Chart*. Este pedido al servidor implica las siguientes tareas: recuperar los tweets/comentarios de la cuenta/noticia seleccionada del período adecuado, identificar las palabras (eliminando *stopwords* y signos de puntuación), y contabilizar la aparición de las palabras relacionadas a cada una de

² Traductor de Google - <https://translate.google.com/>

Tabla 5.1: Ejemplos de palabras asociadas con las ocho emociones primarias.

| | | | | | |
|------------------|--------------|--------------|---------------------|---------------|------------|
| Alegría | felicidad | genial | Tristeza | condolencia | dolor |
| | feliz | maravilloso | | depresión | duelo |
| Confianza | lealtad | ayudar | Disgusto | deshonestidad | suciedad |
| | responsable | agradable | | corrupción | aberración |
| Miedo | asustado | peligroso | Enojo | enfado | queja |
| | cobardía | terror | | molesto | crimen |
| Sorpresa | asombro | sorprendente | Anticipación | intento | ansia |
| | impredecible | incidente | | conjetura | anhelo |

las emociones. Esta visualización muestra los valores acumulados de todo el período de tiempo elegido para cada emoción y es de naturaleza estática.

Por otro lado, la herramienta ofrece un gráfico de burbujas (*Bubble Chart*). En este caso, el servidor recupera los tweets/comentarios y realiza el preprocesamiento de los datos de la misma manera que en el gráfico de radar, pero contabiliza las palabras asociadas a las emociones por día. Esta visualización permite observar la evolución de las emociones a lo largo del tiempo, y ofrece la posibilidad de enfocarse en una o múltiples emociones en simultáneo.

5.4. Diseño

Esta sección presenta las principales decisiones para el diseño del sistema: la Sección 5.4.1) indica la arquitectura del sistema, y la Sección 5.4.2 presenta el modelo de datos.

5.4.1. Diseño de la aplicación

Tal como en la herramienta anterior, el diseño de la estructura se basa en una arquitectura cliente-servidor. El motivo por el que fue elegida esta arquitectura se indica en la Sección 4.4.1. Los componentes de la herramienta pueden observarse en el diagrama de componentes presente en la Figura 5.2. Aquellos presentes en el paquete *Frontend* se ejecutan en el lado del cliente, mientras que el resto se ejecuta en el servidor.

El diseño del sistema tiene tres paquetes principales: *Cliente*, *Cliente* y *Base de datos*. El *Cliente* provee los componentes que implementan todas las funcionalidades del sistema

con las que el usuario puede interactuar. Los tres componentes principales de la interfaz de usuario son *UI Seguimiento de cuentas* para administrar el seguimiento de cuentas, *UI Recuperación de noticias* para obtener la información de noticias, y *UI Visualizaciones* para elegir y mostrar las diferentes visualizaciones.

Por su lado, el paquete *Servidor* implementa el servidor del sistema. Sus componentes principales son: (1) *Pedidos*, por medio del cual recibe los pedidos del cliente y devuelve la información esperada, (2) *Análisis*, encargado de recuperar y analizar los datos requeridos apropiadamente, (3) *Seguimiento de cuentas*, que administrará la información de las cuentas seguidas de modo de guardar los datos correspondientes a cuentas y nuevos posts de manera adecuada, y (4) *Recuperación de noticias* (un paquete auxiliar desarrollado por la tesista) encargado de la recuperación y estructuración de las noticias y sus comentarios³. El componente *APIs*, por su parte, se ocupa de las conexiones con las APIs correspondientes. Finalmente, el paquete *Base de datos* se encarga del manejo de la base de datos, incluidas las operaciones de guardar, obtener y borrar elementos.

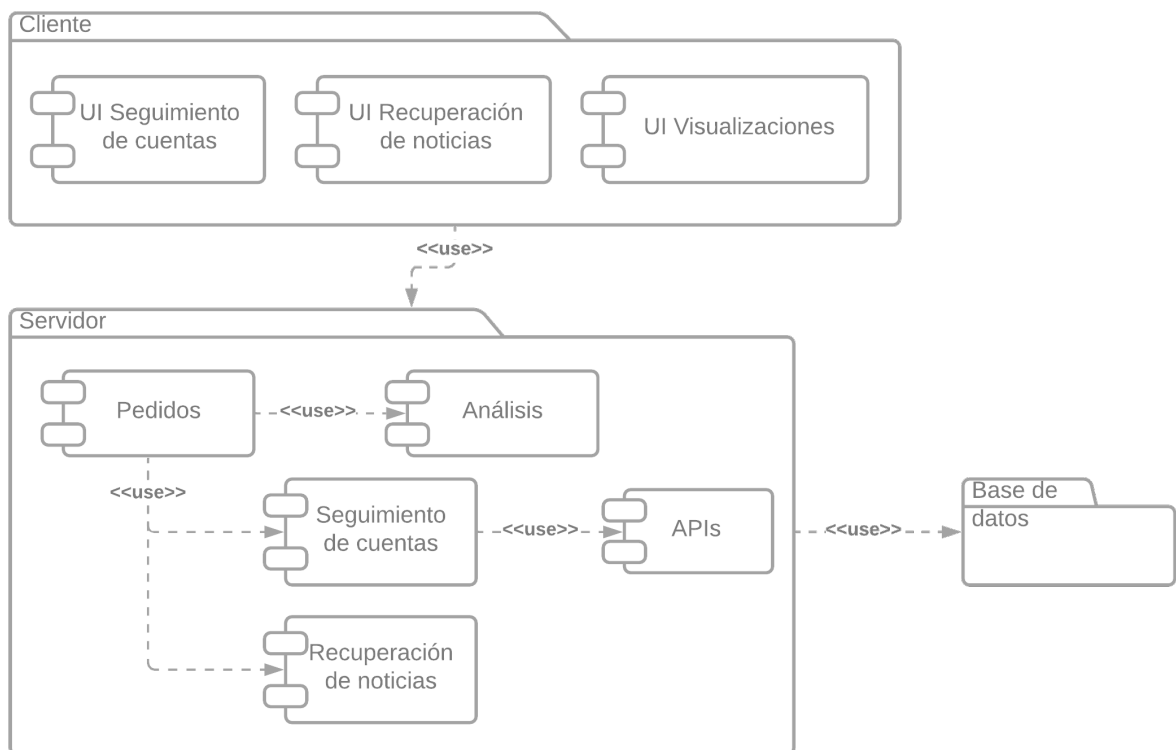


Figura 5.2: Diagrama de componentes.

³ newest - <https://github.com/rbhubert/newest>

5.4.2. Modelo de datos

La herramienta recupera publicaciones, en particular utilizando la API de Twitter y la librería *twarc*⁴. Ambos devuelven información acerca del tweet aparte de su contenido, como puede ser fecha de creación, los recursos multimedia que comparte (enlaces externos, vídeos o imágenes), información del usuario (identificador, nombre de usuario, ubicación, descripción, etc.), si es un *retweet*, el número de favoritos, las diferentes menciones dentro del tweet, el número de amigos, y diferente información acerca de, por ejemplo, la información del perfil (color o imagen de fondo), entre muchos otros.

En la base de datos se guardan los siguientes datos de cada tweet recuperado:

- *id*: número identificador del tweet.
- *user*: información acerca del usuario. En particular:
 - *user_id*: número identificador del usuario.
 - *username*: nombre de usuario.
- *post*: el contenido del tweet. En particular:
 - *original*: el contenido original.
 - *splitted*: una lista de los tokens que conforman el contenido original. Cada elemento de la lista tiene el tipo de token (emoji, hashtag, mención o texto), el valor y las emociones asociadas.
- *creation_time*: el día y horario de creación del tweet.
- *hashtags*: una lista de los hashtags mencionados en el tweet.
- *multimedia*: los recursos multimedia utilizados en el tweet. En particular:
 - *image*: una lista con enlaces a las imágenes.
 - *videos*: una lista con enlaces a los vídeos.
 - *link*: una lista con los enlaces externos.
- *interactions*: las interacciones en respuesta a ese tweet. En particular:
 - *comments*: una lista con los comentarios realizados. Cada comentario es un tweet, por lo tanto tiene los mismos datos aquí listados.
 - *retweets*: el número de *retweets* del tweet.
 - *favorites*: el número de favoritos del tweet.
- *update*: un campo *flag* para indicar si las interacciones relacionadas al tweet han sido actualizadas. Sus valores pueden ser *verdadero* o *falso*.

⁴ twarc - <https://github.com/DocNow/twarc>

La herramienta también permite la recuperación de noticias. En su colección, se guardará la siguiente información de cada noticia:

- *url*: la dirección URL de la noticia.
- *title*: el título de la noticia.
- *content_structured*: el contenido de la noticia de forma estructurada. Se guardan los siguientes datos por cada sección:
 - *section_title*: título de la sección.
 - *text*: el contenido de la sección.
 - *images*: una lista de las imágenes de la sección, donde cada elemento es la dirección URL de la imagen.
 - *links*: una lista con los enlaces de la sección, donde cada elemento será un par *texto del enlace* y *URL del enlace*.
- *content_text*: el contenido de la noticia sin estructurar, solo el texto.
- *creation_time*: el día y horario de creación de la noticia.
- *comments*: una lista con los comentarios realizados. De cada comentario se guarda la siguiente información:
 - *id*: el número identificador del comentario.
 - *text*: el contenido.
 - *username*: el nombre de usuario de quien realizó el comentario.
 - *creation_time*: el día y horario de creación del comentario.
 - *likes*: el número de votos positivos del comentario.
 - *dislikes*: el número de votos negativos del comentario.
 - *replies*: una lista de respuestas al comentario. Cada una de estas respuestas tiene la misma información que un comentario.
- *last_comment*: el identificador del último comentario recuperado.

El detalle sobre la recuperación de los datos, el tiempo de análisis y las cuentas seguidas que se utilizaron para la validación de la herramienta se indicarán en la Sección 5.6 y en la Sección 5.7, correspondientes a los casos de estudio. En ambos casos solo se recupera información de Twitter.

5.5. Implementación y testeo

Para el desarrollo del *frontend* se utilizó el framework Angular⁵ (7.2.10) y Angular Material⁶ (7.3.5) como librería de estilo. Para generar las visualizaciones se usaron las librerías angular-plotly⁷ (1.1.0), D3⁸ (5.9.2), angular4-word-cloud⁹ (1.0.4) y venn.js¹⁰ (0.2.20).

Por otro lado, el *backend* se implementó en Python¹¹ (3.8.1) utilizando flask¹² (1.1.1) para crear la aplicación web. Las librerías más relevantes utilizadas son: tweepy¹³ (3.8.0) para la conexión con la API de Twitter, twarc¹⁴ (1.7.5) para la recuperación de respuestas a tweets, y pandas¹⁵ (0.25.0) y nltk¹⁶ (3.4.4) para la manipulación y análisis de los tweets. Además, se utilizó mongoDB¹⁷ para las bases de datos, haciendo uso de la librería pymongo¹⁸ (3.8.0) para establecer la conexión.

Los testeos se realizaron a nivel de pruebas unitarias por la tesista, y las pruebas de integración y de uso fueron verificadas por las directoras. Así mismo, se complementó el testeo con demos realizadas ante partes interesadas en el desarrollo de la plataforma.

5.6. Caso de estudio: Ministerios de países de Latinoamérica

La herramienta se valida examinando las interacciones gobierno-ciudadano en Twitter. Para construir el conjunto de datos para el análisis del primer caso de estudio, se seleccionaron cinco países hispanos en América Latina: Argentina, Chile, Colombia, México y

⁵ Angular - <https://angular.io/>
⁶ Angular Material - <https://material.angular.io/>
⁷ angular-plotly - <https://github.com/plotly/angular-plotly.js>
⁸ D3 - <https://github.com/d3/d3>
⁹ angular4-word-cloud - <https://github.com/alhazmy13/Angular4-word-cloud>
¹⁰ venn.js - <https://github.com/benfred/venn.js/>
¹¹ Python - <https://www.python.org/>
¹² flask - <https://flask.palletsprojects.com/en/1.1.x/>
¹³ tweepy - <https://www.tweepy.org/>
¹⁴ twarc - <https://github.com/DocNow/twarc>
¹⁵ pandas - <https://pandas.pydata.org/>
¹⁶ nltk - <https://www.nltk.org/>
¹⁷ mongoDB - <https://www.mongodb.com/es>
¹⁸ pymongo - <https://pymongo.readthedocs.io/>

Uruguay, en base a las posiciones de liderazgo que ocupaban en la región en el Índice de Participación Electrónica definido por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas (UNDESA) (UNE20). También se seleccionaron cinco sectores: salud, educación, desarrollo social, trabajo y medio ambiente, ya que estos son los principales sectores gubernamentales evaluados por la Encuesta de Gobierno Electrónico de las Naciones Unidas para medir el Índice de Servicios en Línea. Para cada país y sector se determinó el nombre y cuenta oficial de Twitter del ministerio o secretaría responsable, indicados en la Tabla 5.2. En esta tabla también se indica, junto al país y entre paréntesis, su puesto según el Índice de e-Participación.

El conjunto de datos completo consta de todos los tweets publicados desde las cuentas oficiales de los ministerios o secretarías seleccionados durante el mes de octubre de 2017. También se recopilaron los comentarios hechos por los ciudadanos en respuesta a esos tweets. El conjunto de datos completo incluye 7194 tweets de los ministerios/secretarías y 5038 de ciudadanos.

Tabla 5.2: Los ministerios y las cuentas oficiales de Twitter para cada país y sector analizado. El puesto de cada país según el Índice de e-Participación se indica entre paréntesis.

| Sectores | Países | | | | |
|-------------------|--|--|--|---|---|
| | Argentina (60) | Chile (32) | Colombia (27) | México (14) | Uruguay (39) |
| Salud | Ministerio de Salud | Ministerio de Salud | Ministerio de Salud y de la Protección Social | Secretaría de Salud | Ministerio de la Salud Pública |
| | Msalnacion | Ministeriosalud | MinSaludCol | SSalud_mx | MSPUruguay |
| Desarrollo Social | Ministerio de Desarrollo Social | Ministerio de Desarrollo Social | Ministerio de la Salud y de la Protección Social | Secretaría de Desarrollo Social | Ministerio de Desarrollo Social |
| | MDSNacion | dsocial.gob | MinSaludCol | SEDESOL_mx | Midesuy |
| Educación | Ministerio de Educación | Ministerio de Educación | Ministerio de Educación Nacional | Secretaría de Educación Pública | Ministerio de Educación y Cultura |
| | EducacionAR | Mineduc | Mineducacion | SEP_mx | MEC_Uruguay |
| Trabajo | Ministerio de Trabajo, Empleo y Seguridad Social | Ministerio de Trabajo y Previsión Social | Ministerio de Trabajo | Secretaría de Trabajo y Previsión Social | Ministerio de Trabajo y Seguridad Social |
| | Mintrabajo | MintrabChile | MintrabajoCol | STPS_mx | MTSSuy |
| Medio Ambiente | Ministerio de Ambiente y Desarrollo Sustentable | Ministerio de Medio Ambiente | Ministerio de Ambiente y Desarrollo Sostenible | Secretaría de Medio Ambiente y Recursos Naturales | Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente |
| | Minisambiente | MMACHile | MinAmbienteCo | SEMARNAT_mx | Mvotma_Uruguay |

5.6.1. Análisis de actividad

La Figura 5.3 ofrece una representación gráfica de los niveles de actividad en Twitter de cada uno de los cinco sectores en los países analizados¹⁹. Además de evaluar el número total de publicaciones por institución, este análisis ayuda a identificar eventos destacados que ocurrieron en estos países en días particulares del mes de octubre del 2017. Los niveles máximos de actividad pueden observarse para los Ministerios de Salud, Desarrollo Social y Educación y Cultura de Uruguay. El análisis de los tweets publicados por estos ministerios identifica dos eventos importantes que podrían haber sido la fuente de los picos de actividad, a saber, la «Conferencia Mundial de la Organización Mundial de la Salud sobre Enfermedades No Transmisibles», que tuvo lugar en Montevideo del 18 al 20 de octubre, y la «Segunda Reunión de la Conferencia Regional sobre Desarrollo Social de América Latina y el Caribe», que se realizó en Montevideo del 25 al 27 de octubre.

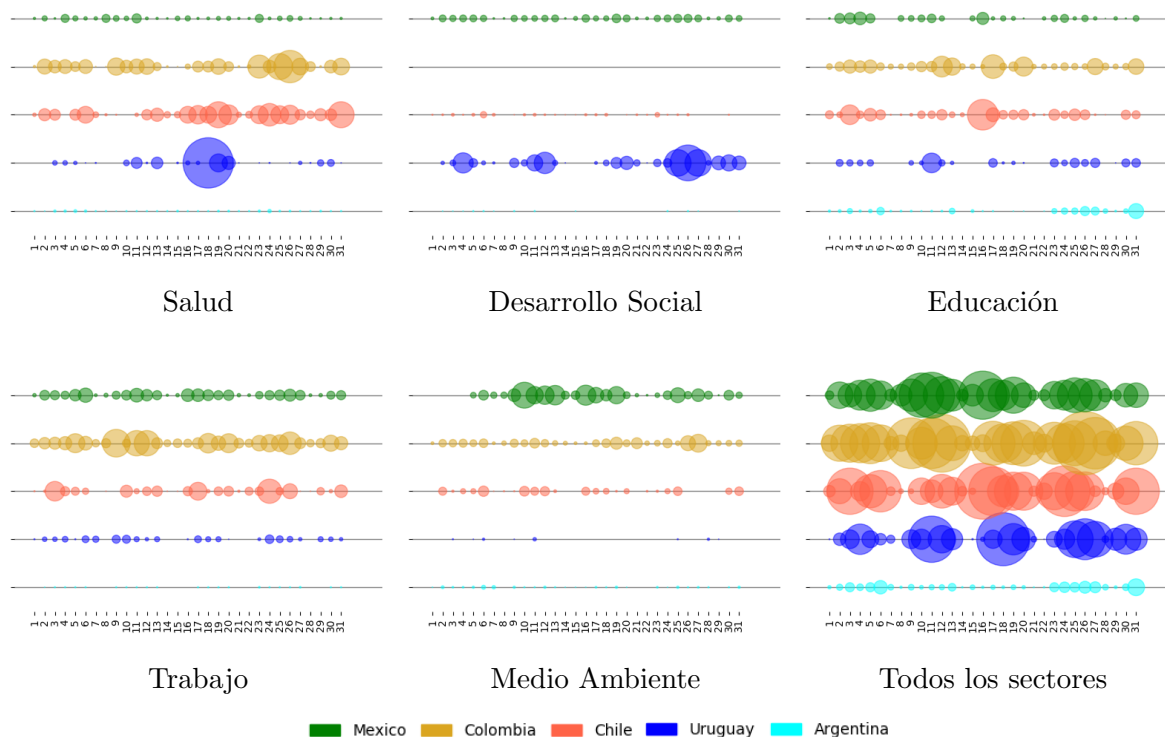


Figura 5.3: Diagramas *event drop* que representa los niveles de actividad de cada sector y de todos los sectores combinados.

¹⁹ Las visualizaciones adicionales pueden encontrarse en <http://ir.cs.uns.edu.ar/publications/downloadextra/84.pdf>

También se pudieron identificar otros eventos ocurridos durante octubre del 2017. Por ejemplo, la Secretaría de Medio Ambiente y Recursos Naturales de México anunció a principios de octubre de 2017 su intención de firmar el «Decreto del Parque Nacional Revillagigedo» mediante el cual el gobierno se compromete con la conservación del Archipiélago de Revillagigedo. Tras el examen de los tweets publicados, los altos niveles de actividad registrados por el Ministerio de Salud y de la Protección Social de Colombia durante los días 23 al 27 de octubre parecen enfocarse en su mayoría a prevenir enfermedades transmitidas por mosquitos, usando el hashtag *#ContraLosMosquitos*. El Ministerio de Educación de Chile también registra un pico de actividad durante el «Día del Profesor», celebrado el 16 de octubre, y el Ministerio de Educación de Argentina durante una evaluación a nivel nacional de futuros docentes el 31 de octubre.

Además de los diagramas para cada sector, la Figura 5.3 también muestra la visualización de los niveles totales de actividad de cada país después de agregar todas las actividades específicas del sector. El orden del país más activo al menos activo, considerando el número de publicaciones publicadas es: Colombia (2244), Chile (1729), México (1721), Uruguay (1250) y Argentina (406).

5.6.2. Identificación de temas destacados

La Figura 5.4 muestra las nubes de palabras asociadas con los tweets publicados por los Ministerios/Secretarías¹⁹ de Desarrollo Social de México, Chile, Uruguay y Argentina. Colombia no tiene un Ministerio dedicado al Desarrollo Social. El análisis de estas nubes de palabras permite identificar los principales temas o eventos tratados por este sector en cada país durante octubre del 2017. En particular, uno de los términos más utilizados por la Secretaría de Desarrollo Social de México fue *sismo*, reflejando la ocurrencia de los terremotos que azotaron México en septiembre del 2017. El término *Chiapas* es también prominente dados los importantes daños a la infraestructura producidos por el terremoto en el estado de Chiapas y las acciones tomadas por la Secretaría luego del evento. El hashtag *#FuerzaMéxico* también se utilizó con frecuencia en este contexto.

A diferencia de México, la Secretaría de Desarrollo Social de Chile no se centró en temas específicos durante el período analizado. Sus anuncios incluyen menciones al Ministro de Desarrollo Social de Chile (Marcos Barraza) y al proceso conocido como «Proceso Constituyente Indígena», que tiene como objetivo permitir que las comunidades indíge-

nas y tribales participen en los procesos de toma de decisiones que afectan sus vidas. En Uruguay, los anuncios hechos por el Ministerio de Desarrollo Social ponen especial énfasis en las palabras *mujeres* y *jóvenes*. Asimismo, el Ministerio de Desarrollo Social argentino también hizo referencias a los jóvenes, agregando además temas como *programas*, *actividades* y *talleres*. También hace referencia al hashtag *#ElEstadoEnTuBarrio*, un programa que facilita el acceso a servicios y beneficios a los ciudadanos de sus barrios.



Figura 5.4: Nubes de palabras destacando los principales temas tratados por los Ministerios/Secretarías de Desarrollo Social.

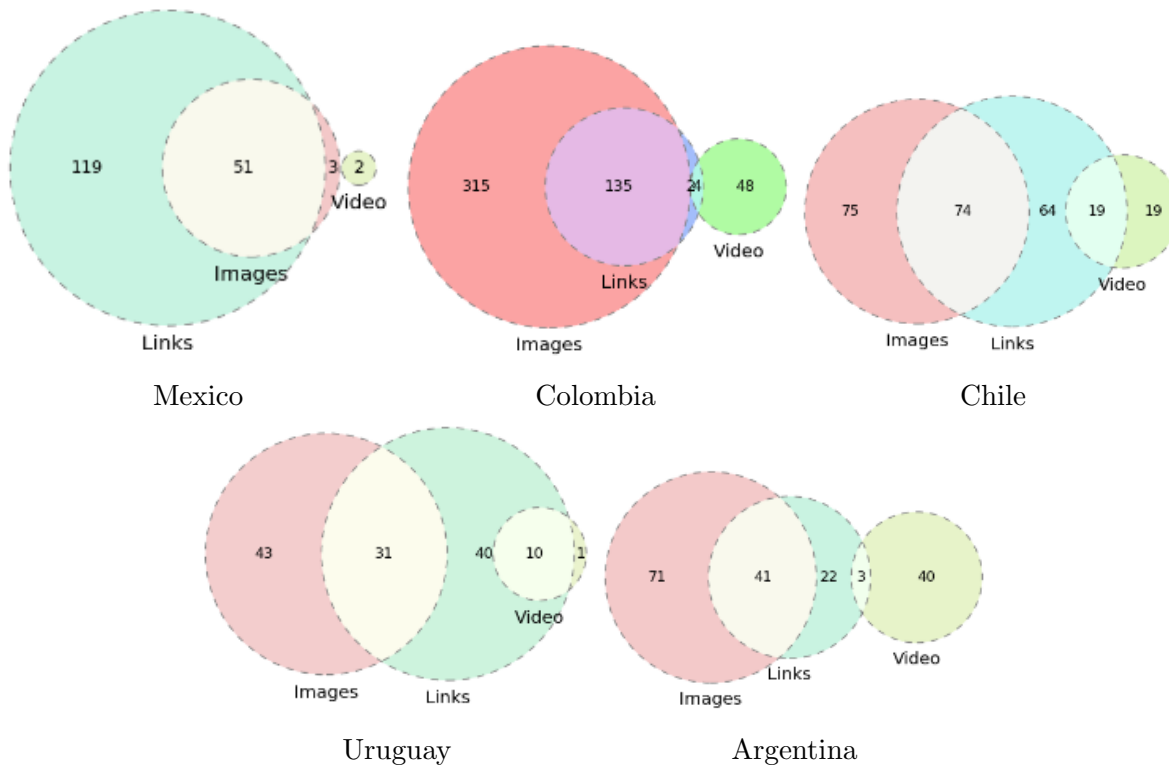


Figura 5.5: Diagramas de Euler coloreados mostrando el número de enlaces, imágenes y vídeos compartidos por los Ministerios/Secretarías de Educación de cada país.

5.6.3. Análisis del uso de recursos multimedia

En la Figura 5.5, los diagramas de Euler¹⁹ coloreados ayudan a reconocer las estrategias de uso de recursos multimedia más comunes adoptadas por el sector educativo de los cinco países durante octubre de 2017. Por ejemplo, México publicó 119 tweets que contienen enlaces, tres con imágenes y dos con vídeos. También publicó 51 tweets que contienen enlaces e imágenes. Se puede observar que México comparte principalmente enlaces, mientras que Colombia comparte principalmente imágenes.

De manera más general, la Figura 5.6 muestra la cantidad de recursos de cada tipo compartidos por cada sector en los cinco países. En base a este análisis, se puede concluir que los sectores de trabajo y salud son los que más utilizan recursos multimedia, siendo las imágenes el tipo de recurso más común.

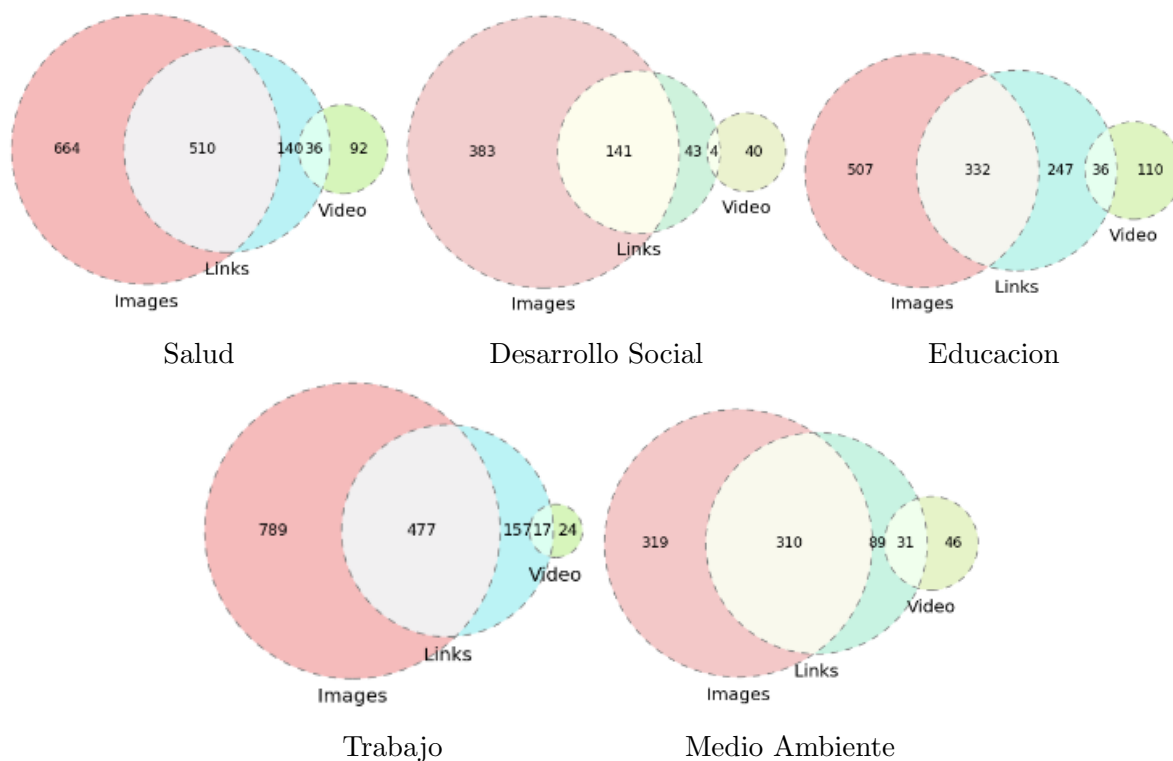


Figura 5.6: Diagramas de Euler coloreados mostrando el número de enlaces, imágenes y vídeos compartidos por cada sector en los cinco países.

5.6.4. Análisis de las respuestas ciudadanas

La Figura 5.7 presenta un diagrama de Hasse extendido¹⁹ que representa la cantidad de comentarios, favoritos y *retweets* publicados por los ciudadanos en respuesta a los anuncios realizados por las Secretarías/Ministerios de Trabajo en cada uno de los países analizados. En general, la cantidad de comentarios es menor que la cantidad de *retweets* o favoritos. En la mayoría de los casos, los tweets marcados como favoritos también se comparten. Cabe mencionar que en este sector, Colombia tiene el nivel más alto de participación, mientras que Argentina tiene el más bajo. Una comparación entre los niveles de actividad del sector de trabajo en la Figura 5.3 y el grado de respuesta ciudadana en la Figura 5.7 indica que los países con los niveles más altos de actividad (Colombia, México y Chile) son también los que presentan los niveles más altos de interacción ciudadana.

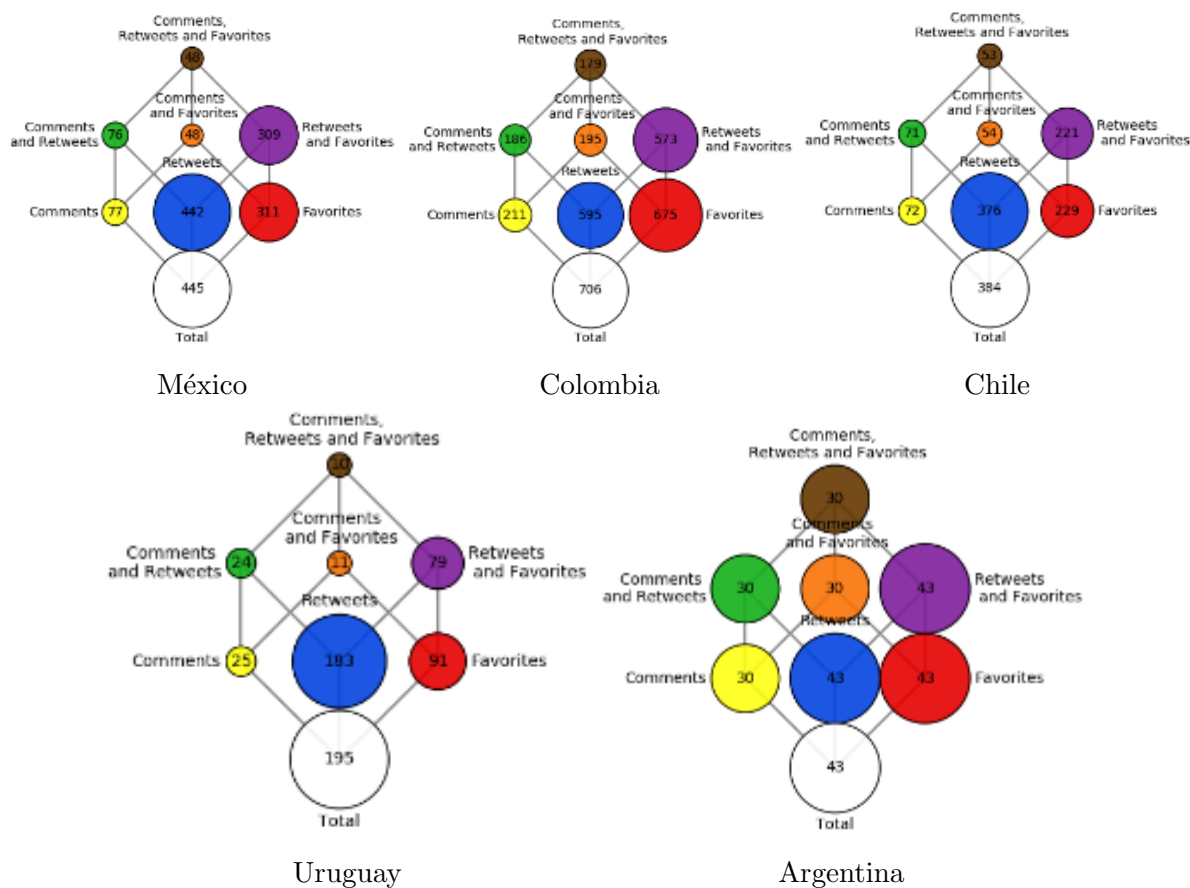


Figura 5.7: Diagramas de Hasse extendidos que representan la respuesta de los ciudadanos a través de comentarios, *retweets* y favoritos de tweets de Ministerios/Secretarías de Trabajo de cada país.

De manera más general, la Figura 5.8 presenta el número total de comentarios, favoritos y *retweets* publicados por los ciudadanos en respuesta a los tweets realizados por los cinco ministerios/secretarías en cada uno de los cinco países. El análisis estableció una correspondencia entre los niveles de actividad del gobierno en Twitter y el grado de respuesta ciudadana. Además, los países con mayor índice de e-Participación según UNDESA son también los que reportan mayores niveles de actividad en Twitter durante el período analizado. Sin embargo, esta correspondencia no es exacta. Por ejemplo, Colombia, que en el momento del análisis ocupaba el puesto 27 en el índice de e-Participación, reportaba un nivel de actividad más alto que México, que ocupaba el puesto 14.

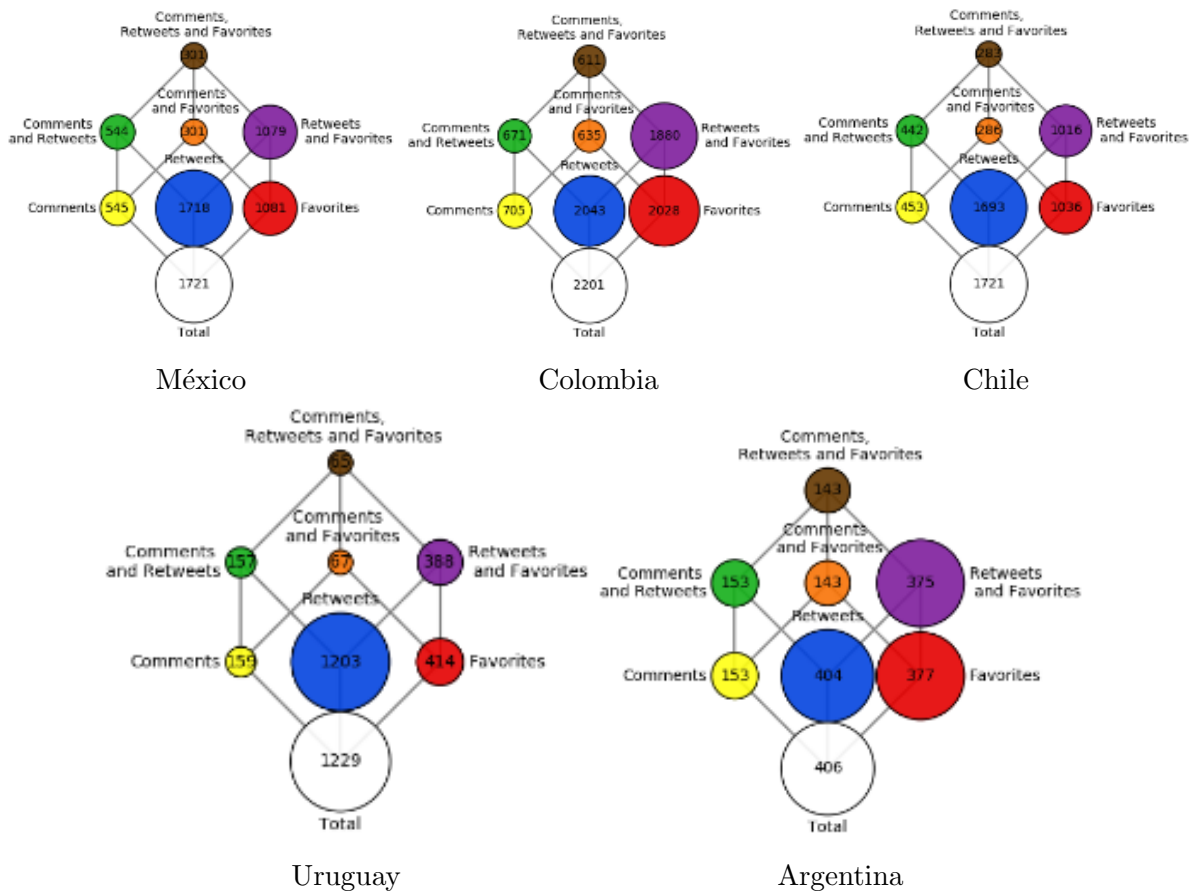


Figura 5.8: Diagramas de Hasse extendidos que representan la respuesta de los ciudadanos a través de comentarios, *retweets* y favoritos de tweets de todos los Ministerios/Secretarías de cada país.

5.6.5. Análisis de sentimientos

Este análisis se aplicó para evaluar el estado de ánimo general de los ciudadanos de un país en respuesta a los tweets realizados por ministerios/secretarías gubernamentales particulares. Luego de aplicar esta técnica a los comentarios publicados por los ciudadanos, se generaron representaciones gráficas para visualizar las emociones predominantes encontradas en los comentarios analizados. La Figura 5.9 presenta las visualizaciones¹⁹ resultantes para el sector de salud. De los gráficos se puede concluir que la *confianza* es la emoción más común para los tweets que se originan en el sector de salud, mientras que en algunos países como Argentina o Uruguay el *miedo* es casi igual de predominante.

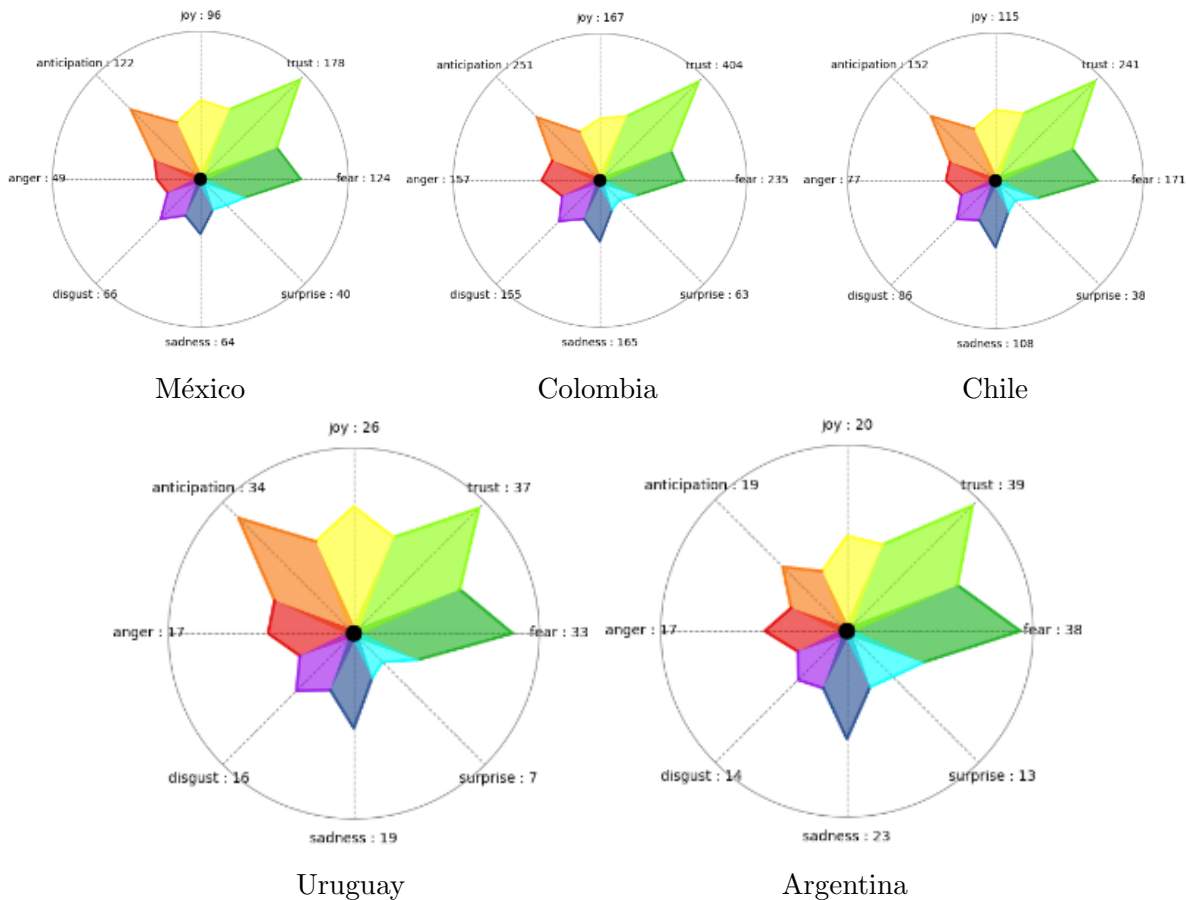


Figura 5.9: Gráfico de radar mostrando las emociones identificadas en las respuestas ciudadanas a las publicaciones realizadas por los Ministerios/Secretarías de Salud en los cinco países analizados.

5.6.6. Análisis combinado

Para ilustrar cómo se pueden combinar diferentes tipos de análisis, examinamos las posibles correlaciones entre la cantidad de imágenes, vídeos, enlaces y tweets publicados por los ministerios argentinos y el nivel de respuesta ciudadana mediante comentarios, *retweets* y favoritos. La Figura 5.10 muestra, mediante un gráfico de líneas con anotaciones¹⁹, una correlación a lo largo de un período de un mes para el Ministerio de Salud de Argentina, e indica que las imágenes, enlaces y tweets generan un mayor nivel de respuesta de los ciudadanos que los vídeos. Las anotaciones del gráfico permiten determinar los temas destacados cuando se observan picos de actividad.

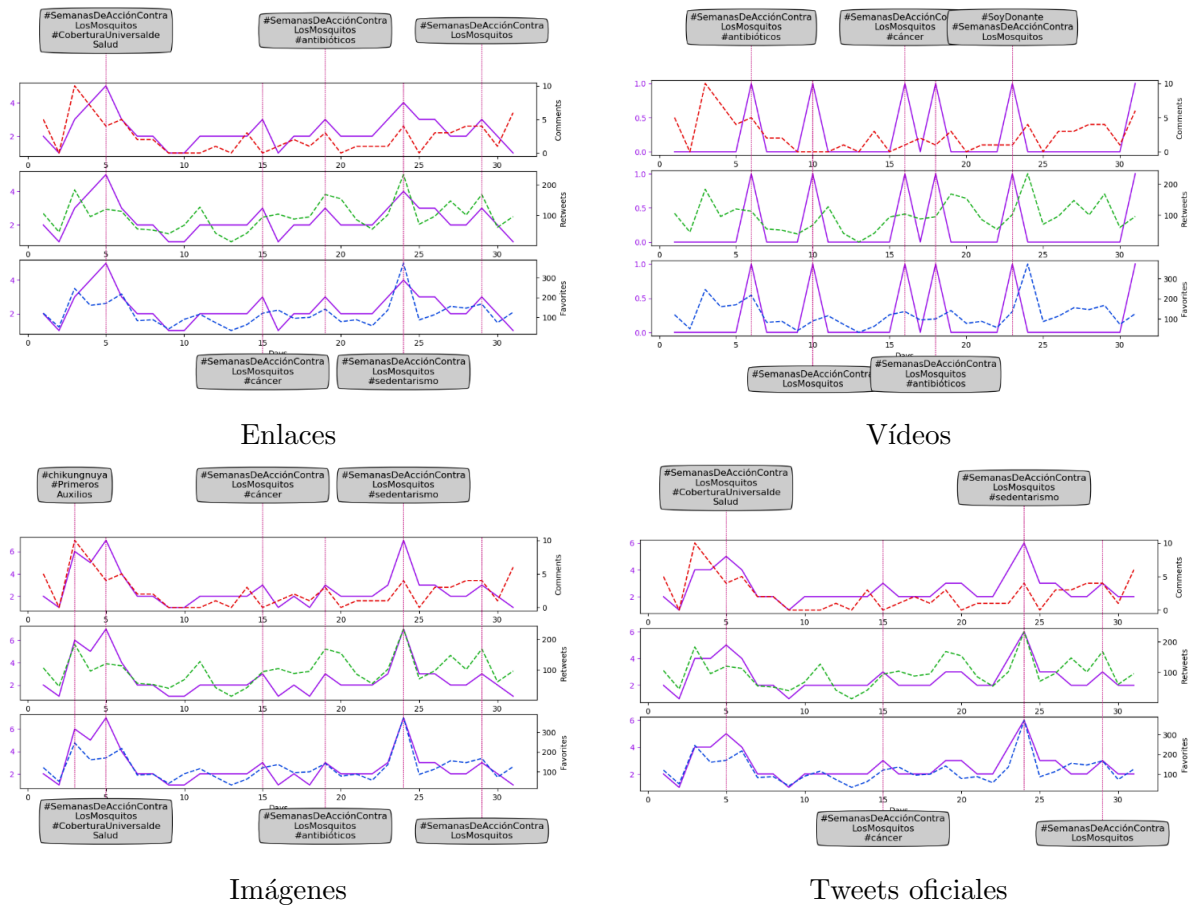


Figura 5.10: Gráfico de líneas con anotaciones que muestran las correlaciones entre enlaces, vídeos, imágenes y tweets publicados por el Ministerio de Salud de Argentina y los comentarios, *retweets* y favoritos de los ciudadanos.

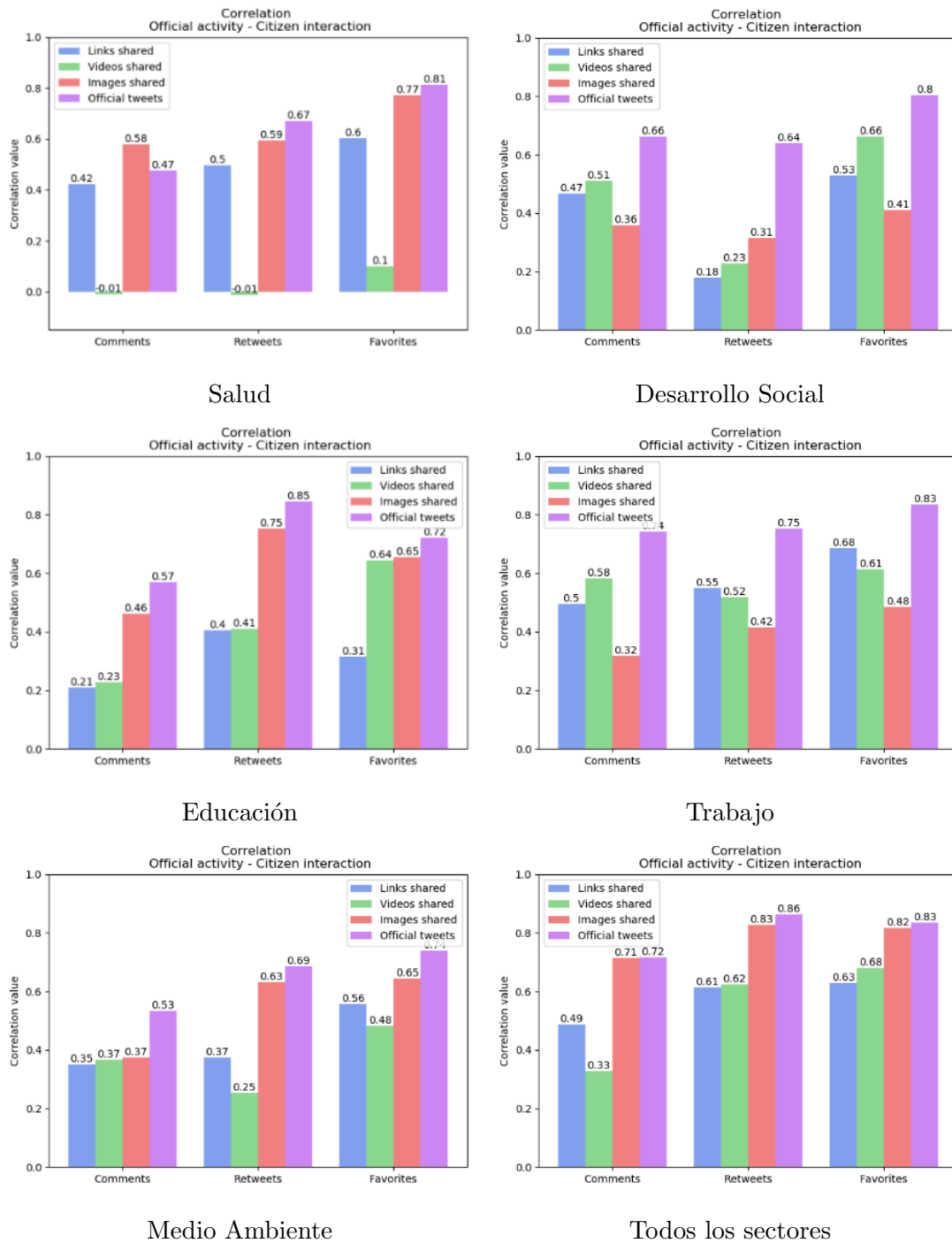


Figura 5.11: Gráficos de barras que muestran correlaciones entre enlaces, vídeos, imágenes y tweets publicados por los cinco ministerios argentinos y comentarios, *retweets* y favoritos expresados por los ciudadanos.

Los gráficos de barras¹⁹ de la Figura 5.11 muestran los valores correlacionados entre los cuatro tipos de actividad oficial (enlaces, vídeos, imágenes y tweets publicados) y los tres tipos de respuestas ciudadanas (comentarios, *retweets* y favoritos) para cinco sectores gubernamentales en Argentina. Para el sector de la salud, existen correlaciones altas entre la respuesta ciudadana y todo tipo de actividad oficial, excepto cuando se comparten vídeos. Otros sectores tienen fuertes correlaciones positivas entre las respuestas ciudadanas y todo tipo de actividad oficial, incluido cuando se comparten vídeos. Sin embargo, también se pueden identificar algunas excepciones sin correlación aparente.

5.7. Caso de estudio: Secretarías de México

La segunda validación se centra en el análisis de la interacción gobierno-ciudadano por medio de Twitter en México. Esta elección está basada en su posición de liderazgo en la región según el Índice de e-Participación de UNDESA. Al igual que el caso de estudio anterior, se seleccionaron cinco sectores gubernamentales: salud, educación, desarrollo social, trabajo y medio ambiente, siendo estos los principales sectores gubernamentales evaluados por la Encuesta de e-Gobierno de las Naciones Unidas para medir el Índice de Servicios en Línea. Para cada sector, se identificó el nombre oficial y la cuenta de Twitter de la secretaría responsable, indicadas en la Tabla 5.2.

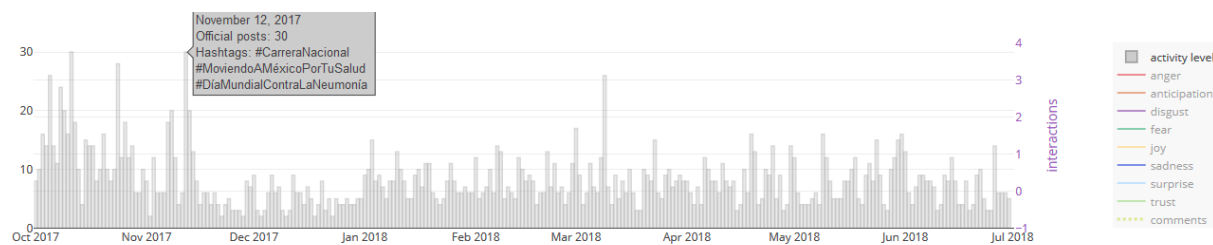
El conjunto de datos completo consta de todos los tweets publicados desde las cuentas oficiales de las secretarías entre octubre del 2017 y junio del 2018, y de los comentarios de los ciudadanos en respuesta a estos. En total, el conjunto de datos completo comprende 30.657 tweets, 14.657 (48 %) de secretarías y 16.000 (52 %) de ciudadanos.

5.7.1. Análisis de actividad

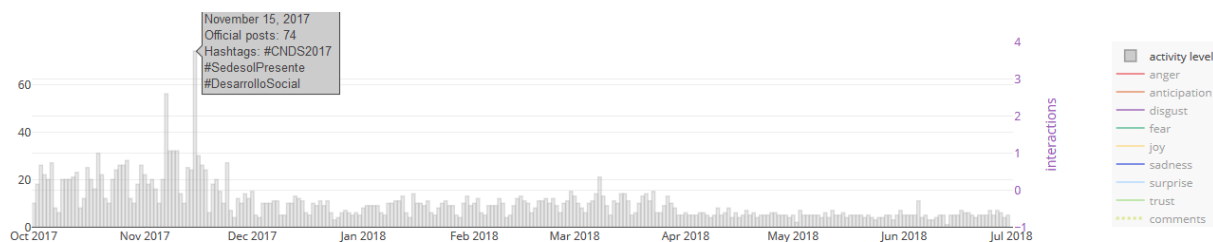
La Figura 5.12 ofrece una visualización de los niveles de actividad de las cuentas oficiales de Twitter de las cinco secretarías elegidas. En el sector de desarrollo social se pueden observar algunos picos de actividad en noviembre del 2017, específicamente el 6 y el 14. Mediante el análisis de los tweets publicados y los hashtags utilizados se identificaron dos eventos importantes causantes de los picos: «Premio Nacional de la Juventud» (*#PNJ2017*), que tuvo lugar el 4 de noviembre, y «Comisión Nacional de Desarrollo Social» (*#CNDS2017*), que se llevó a cabo el 14 de noviembre.

El sector de medio ambiente tuvo un pico de actividad el 7 de noviembre del 2017 asociado a una serie de publicaciones dedicadas a la audiencia parlamentaria (*#comparescencia*). También se pueden observar niveles de actividad altos en el sector educación durante octubre y noviembre del 2017. El pico del 3 de octubre del 2017 corresponde al regreso a clases luego de los terremotos de septiembre del 2017 (*#RegresoAClases*). Asimismo, podemos observar varios picos de actividad en el sector salud, relacionados con los eventos de *#AvancesEnSalud* realizados por la Comisión de Salud el 10 de octubre del 2017, el *#DíaDelMédico* el 23 de octubre del 2017, y el Día Internacional de la Mujer el 8 de marzo del 2018.

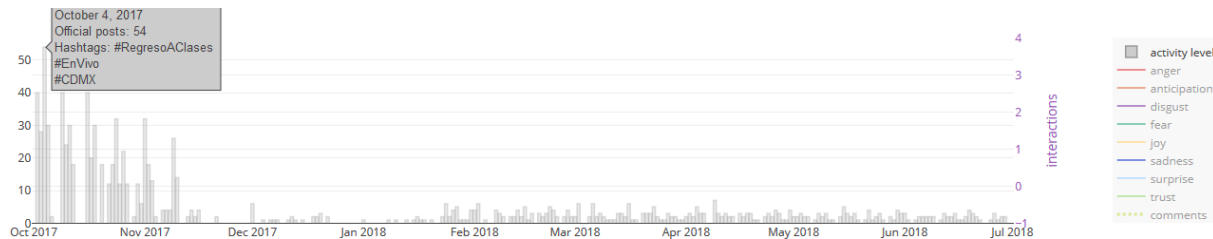
Una situación similar ocurre con el sector laboral, con la mayoría de actividades concentradas en octubre y noviembre del 2017. Durante octubre del 2017 podemos observar



Salud



Desarrollo Social



Educación

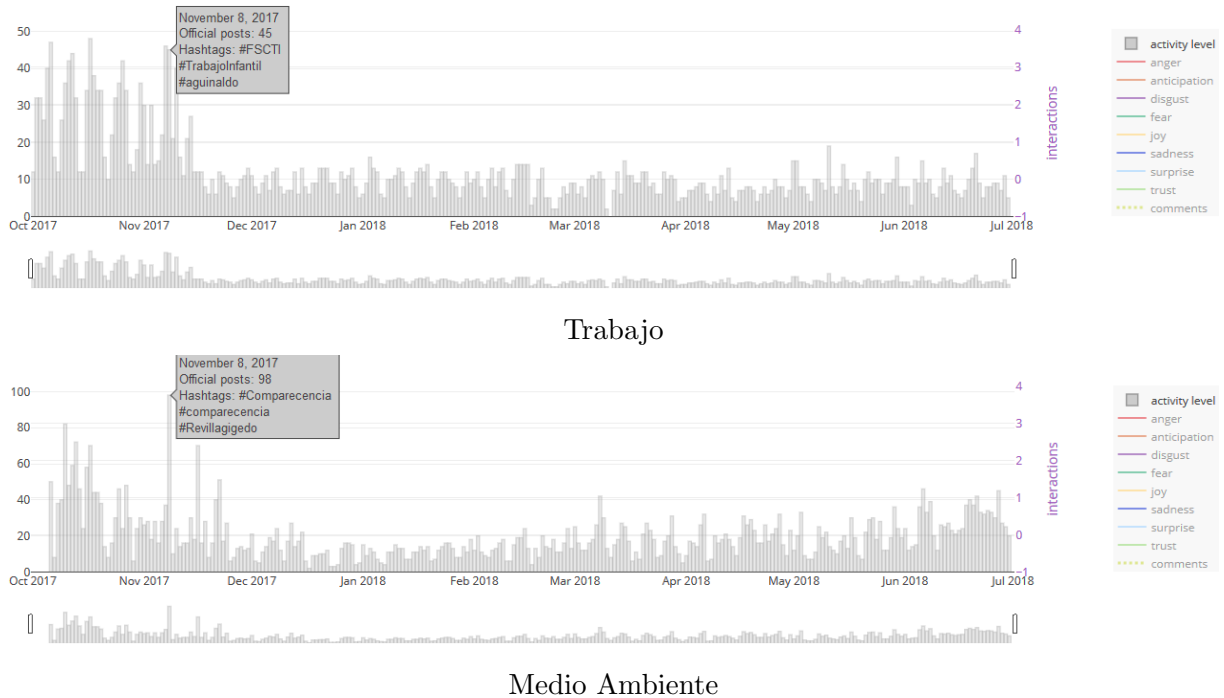


Figura 5.12: Niveles de actividad de los cinco sectores analizados.

tweets relacionados con los terremotos y sus consecuencias en Chiapas (*#sismos*, *#Chiapas*), principalmente el 5 de octubre, y tweets del 16 de octubre del 2017 sobre medidas para prevenir y erradicar el trabajo infantil (*#TrabajoInfantil*). Los tweets sobre trabajo infantil también se pueden observar durante el pico del 6 de noviembre del 2017, cuando el Foro España-Américas (*#FSCTI*) impulsó alianzas entre ONG españolas y latinoamericanas para luchar contra el trabajo infantil y adolescente.

5.7.2. Identificación de temas destacados

La Figura 5.13 muestra las nubes de palabras generadas a partir de los tweets publicados por las cinco secretarías durante noviembre del 2017. En particular, los términos más utilizados por la Secretaría de Desarrollo Social fueron *social* o *sedesol* y los hashtags *#SabíasQue* y *#SedesolPresente*. Un hashtag muy utilizado en la Secretaría de Medio Ambiente y Recursos Naturales fue *#Revillagigedo*, probablemente relacionado con la firma del «Decreto del Parque Nacional Revillagigedo», mediante el cual el Gobierno de México se comprometió con la conservación del Archipiélago de Revillagigedo.

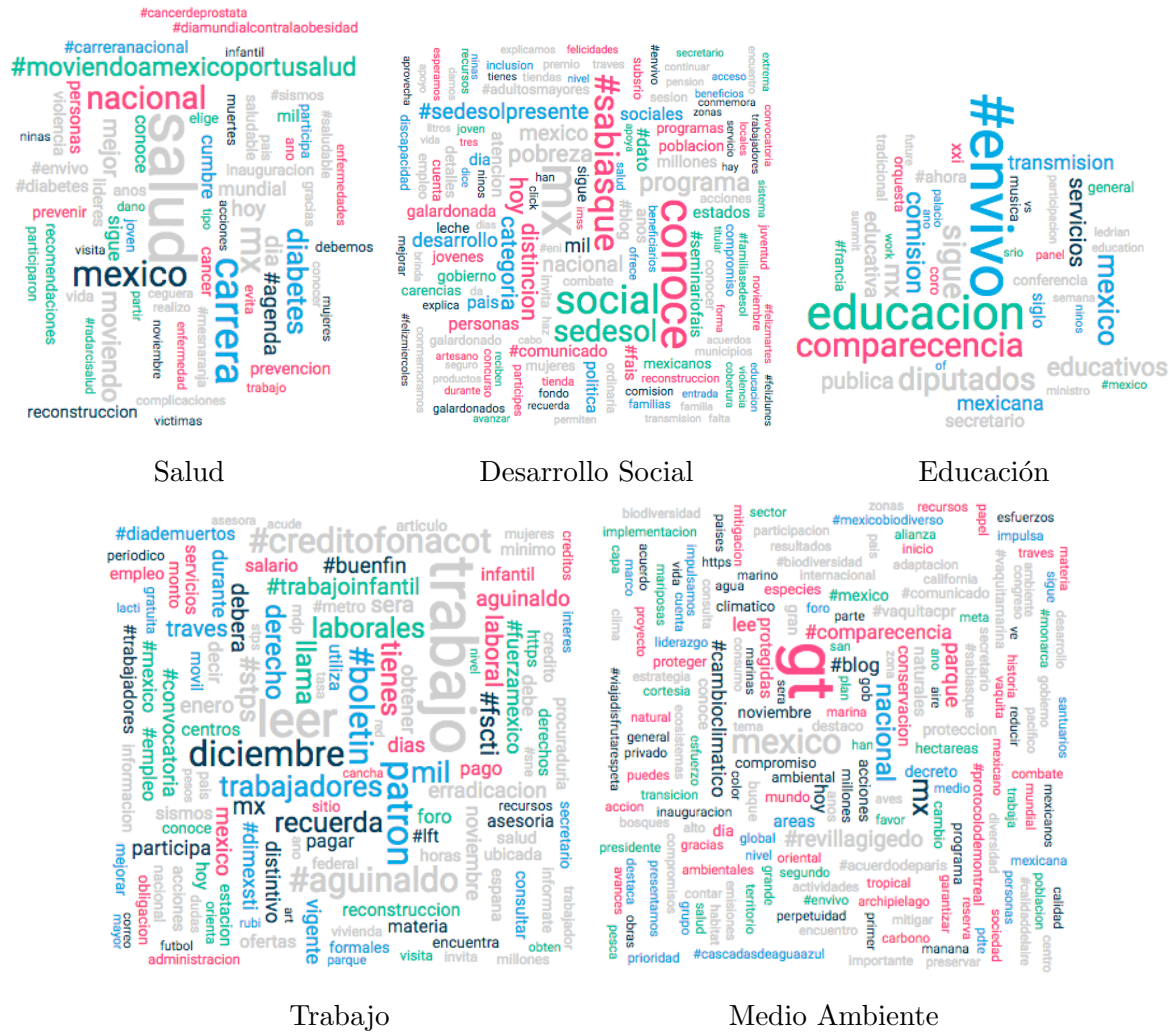


Figura 5.13: Los principales temas tratados por las cinco secretarías analizadas.

Todas las palabras utilizadas en la nube de palabras del sector educativo están relacionadas con la audiencia en vivo (*#envivo*, *#comparecencia*) del Quinto Informe de Gobierno. Las palabras más comunes utilizadas en los tweets de la Secretaría de Salud fueron sobre el proyecto *#MoviendoMexicoPorTuSalud*. Finalmente, entre los tweets relacionados con el trabajo, las palabras más utilizadas incluyen *#trabajo*, *#diciembre*, *#aginaldo* y *#créditoFONACOT* (una institución financiera del gobierno).

5.7.3. Análisis del uso de recursos multimedia

La Figura 5.14 indica las estrategias adoptadas por las cinco secretarías del uso de recursos multimedia en sus comunicaciones durante el mes de enero de 2018. Por ejemplo, la Secretaría de Trabajo publicó 175 tweets con enlaces, 384 con imágenes y 11 con vídeos. Asimismo, publicó 162 tweets que contenían enlaces e imágenes, y solo uno que contenía enlaces y vídeos simultáneamente.

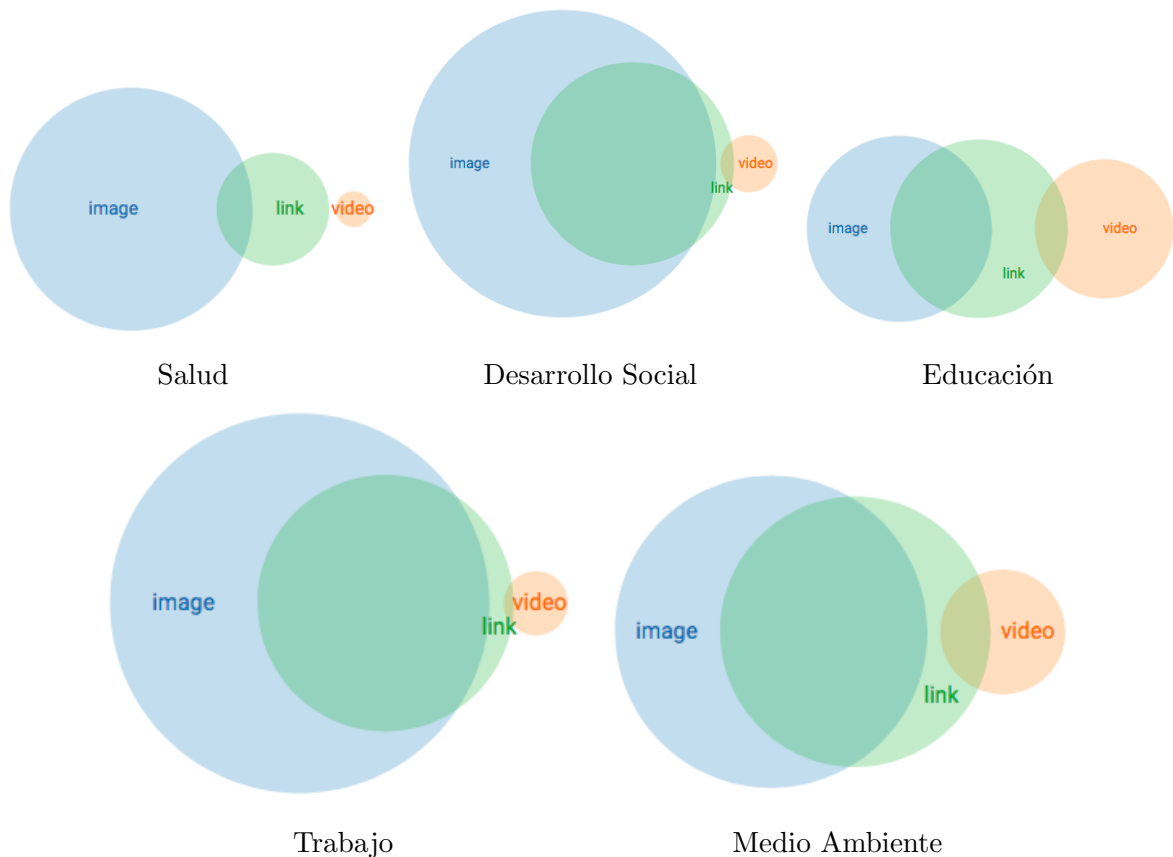


Figura 5.14: Los diferentes tipos de recursos multimedia utilizados por las secretarías analizadas.

Podemos observar que algunas secretarías, como la Secretaría de Desarrollo Social y la Secretaría de Trabajo, utilizan mayoritariamente imágenes para compartir contenidos, mientras que otras comparten imágenes y enlaces. Además, la Secretaría de Educación comparte muchos más vídeos que otros sectores.

5.7.4. Análisis de las respuestas ciudadanas

La Figura 5.15 presenta una visualización del número de comentarios, favoritos y *retweets* de los ciudadanos en respuesta a los anuncios realizados por las cinco secretarías entre octubre de 2017 y junio de 2018. En general, el número de comentarios es menor que el número de *retweets* o favoritos. En la mayoría de los casos, los tweets que se marcan como favoritos también se comparten.

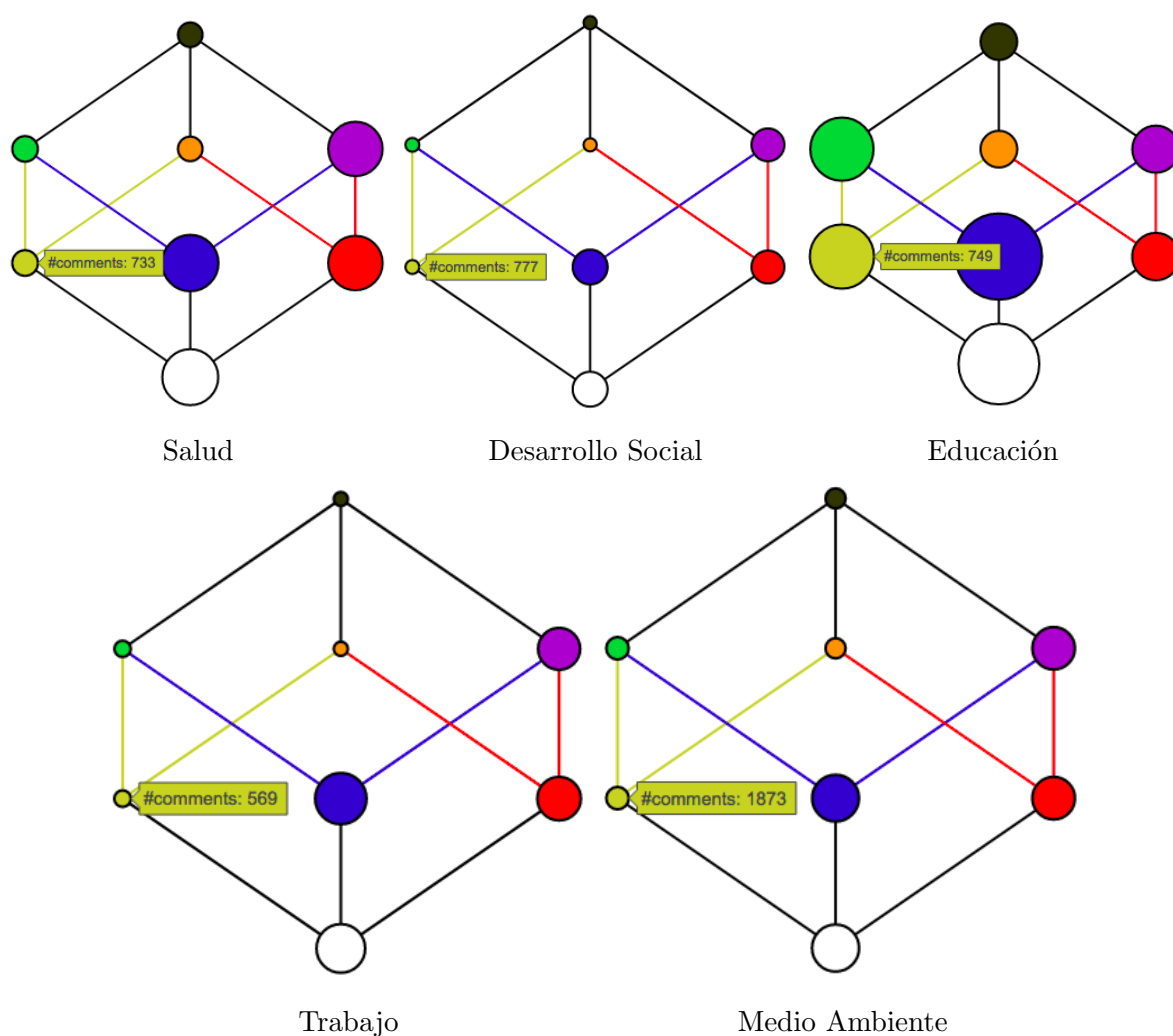


Figura 5.15: Respuestas ciudadanas a través de comentarios, *retweets* y favoritos a los tweets realizados por las cinco secretarías de gobierno durante el período de interés.

La Secretaría de Medio Ambiente tiene el nivel más alto de participación mientras que la Secretaría de Educación tiene el nivel más bajo. Una comparación entre los niveles de actividad de la Figura 5.12 y el grado de respuesta ciudadana en la Figura 5.15 indica

que los sectores gubernamentales con los niveles más altos de actividad, como el de medio ambiente, el de trabajo o el de desarrollo social, son también los que tienen los niveles más altos de participación ciudadana. Además, se puede observar que la Secretaría de Educación tiene más comentarios que la Secretaría de Trabajo y un número similar a la Secretaría de Salud y la Secretaría de Desarrollo Social, aún cuando tiene menores niveles de actividad e interacción que esas secretarías.

La herramienta permite interactuar con las visualizaciones. Por ejemplo, se puede observar los niveles de respuesta ciudadana junto al nivel de actividad (visualización de la Figura 5.12) a lo largo del tiempo en diferentes colores.

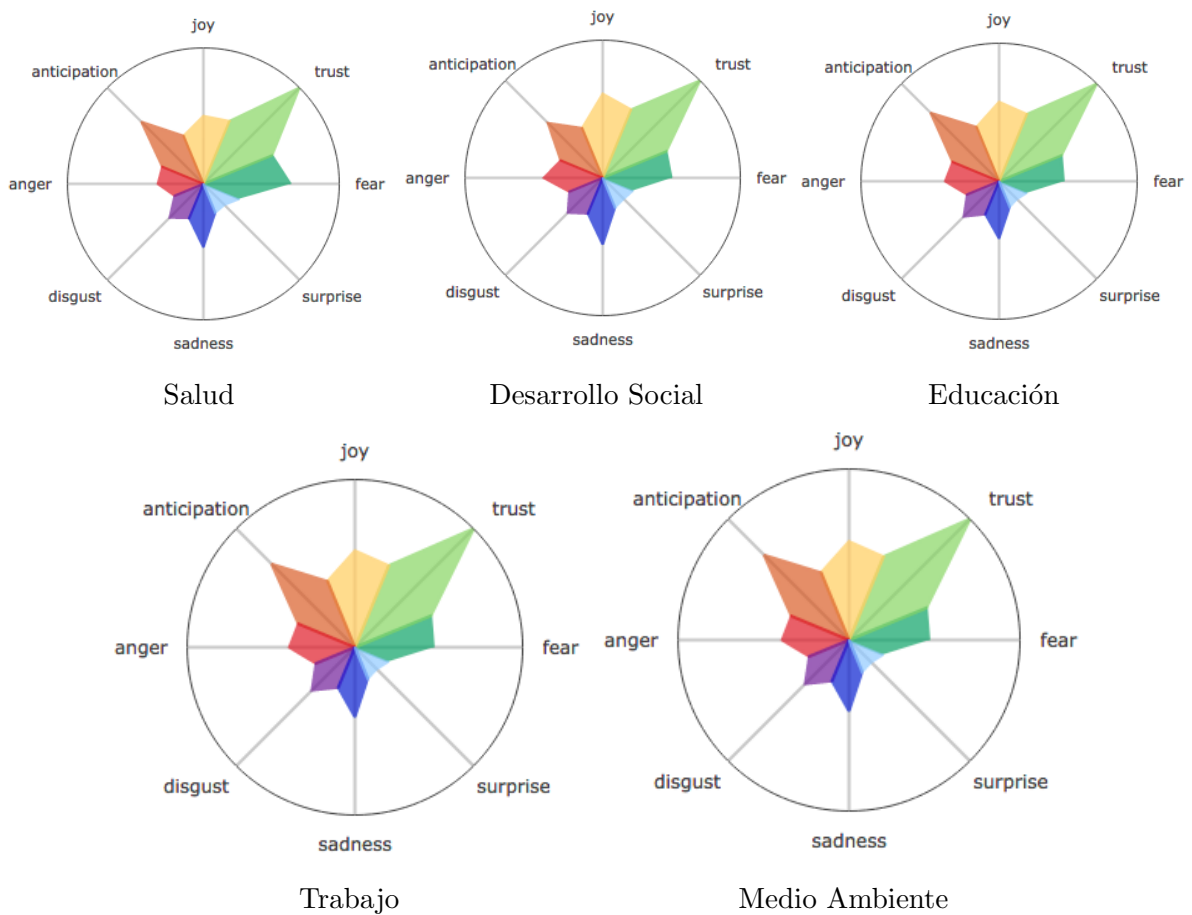
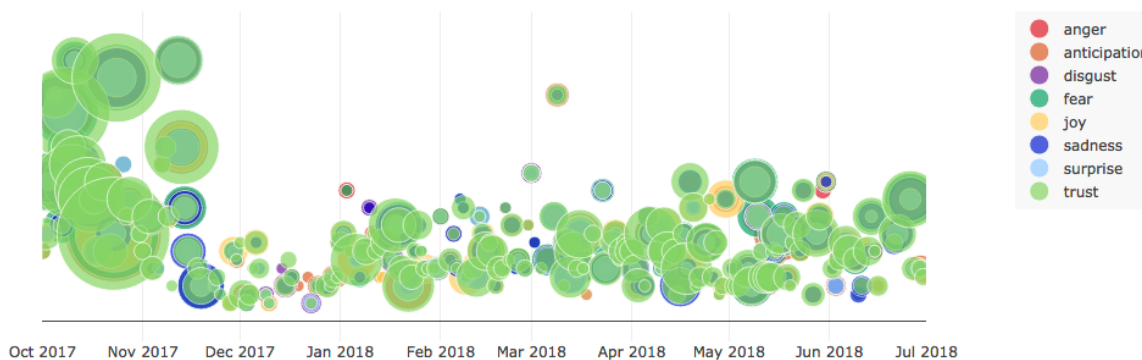


Figura 5.16: Emociones expresadas por los ciudadanos en sus respuestas a los tweets realizados por las cinco secretarías de gobierno.

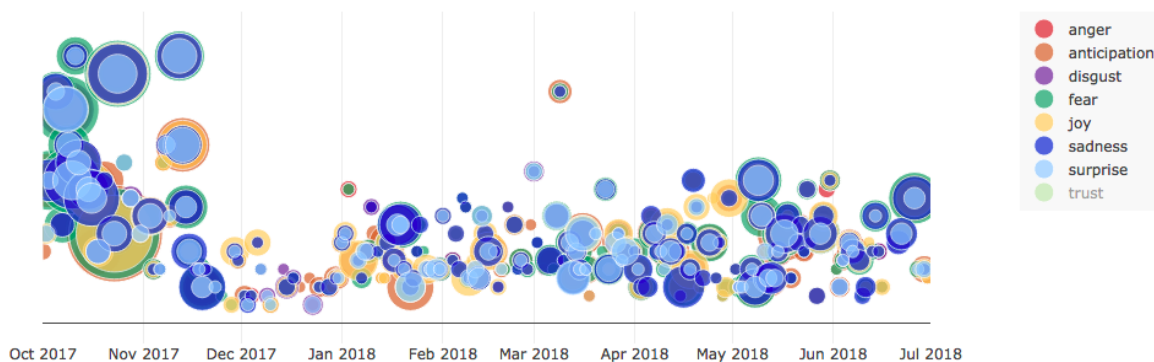
5.7.5. Análisis de sentimientos

La Figura 5.16 presenta las emociones identificadas en los comentarios de los ciudadanos en respuesta a los tweets de las secretarías. Se puede concluir que la confianza es la emoción más común para los cinco sectores, mientras que en algunos sectores, como el de medio ambiente y el de trabajo, la anticipación es casi tan común como la confianza. También podemos notar que el miedo es alto en respuesta a los tweets del sector de salud.

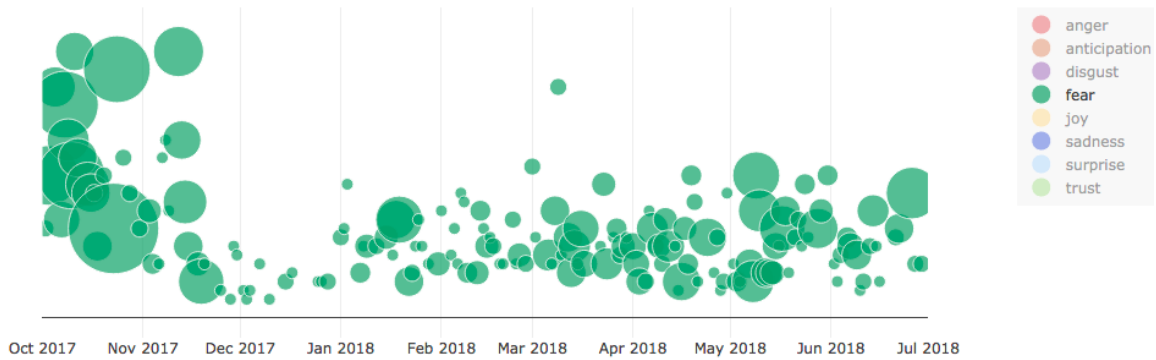
Otra forma de analizar las emociones es mediante los gráficos de burbujas. El nivel de emociones en un día dado está representado por una burbuja: cuanto mayor es la burbuja, mayor es el valor de la emoción. La burbuja se coloca verticalmente dependiendo del número de publicaciones publicadas por la secretaría bajo análisis. Con esta visualización podemos identificar las emociones predominantes expresadas por los ciudadanos en diferentes períodos de tiempo y la evolución de las emociones a lo largo del tiempo. En la Figura 5.17a, podemos identificar niveles altos de confianza, siendo la emoción predominante para el sector de la salud.



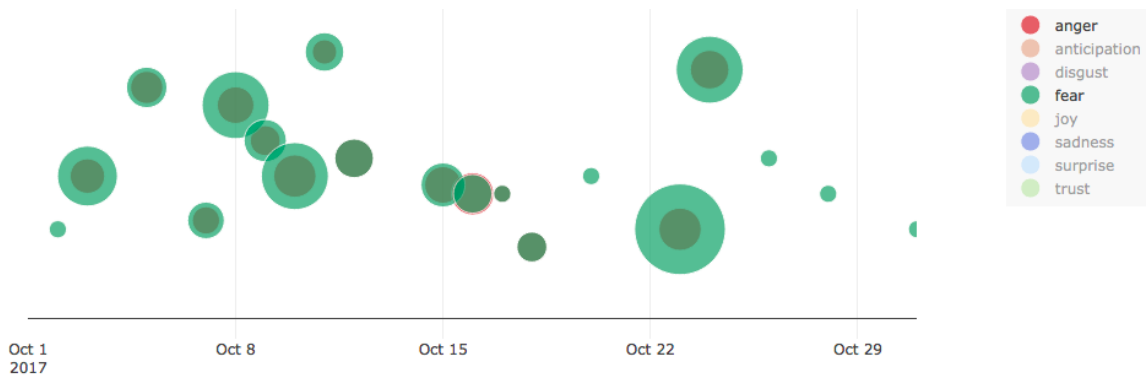
(a) todas las emociones identificadas entre octubre del 2017 y junio del 2018.



(b) todas las emociones excepto *confianza* entre octubre del 2017 y junio del 2018.



(c) evolución de la emoción *miedo* entre octubre del 2017 y junio del 2018.



(d) evolución de las emociones *miedo* e *ira* durante octubre del 2017.

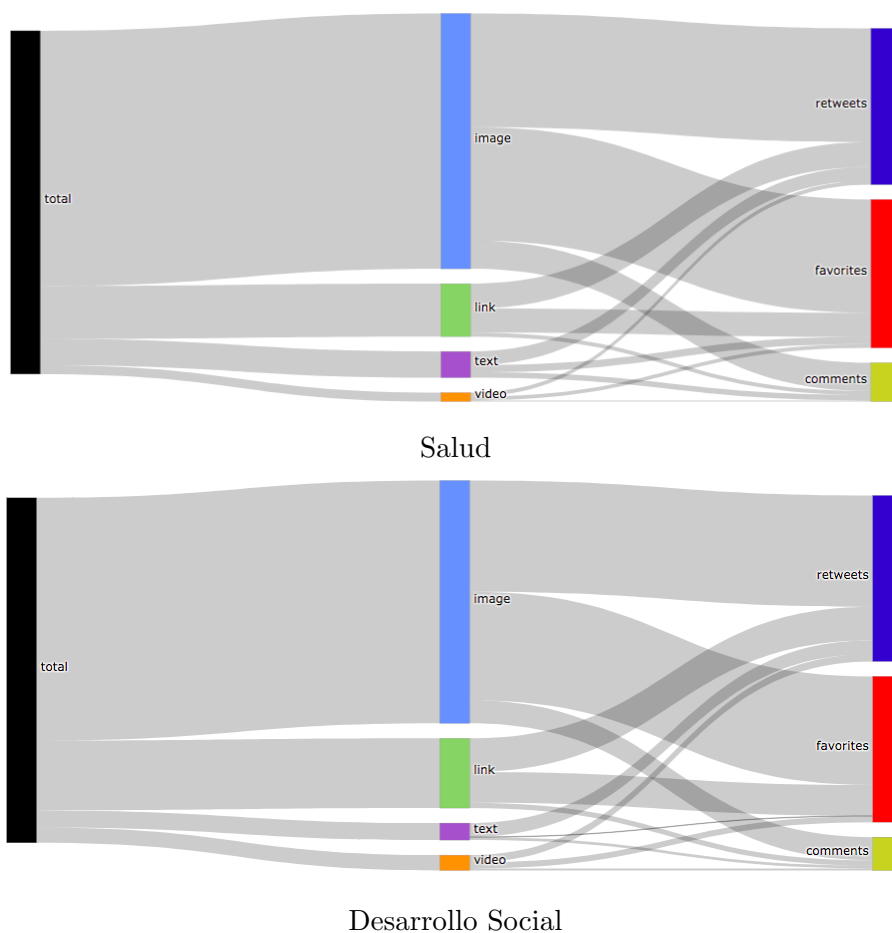
Figura 5.17: Emociones expresadas por la ciudadanía en sus respuestas a los tweets realizados por la Secretaría de Salud y su evolución durante el período de análisis.

La visualización permite deseleccionar algunas emociones, por lo que deseccionamos la confianza para observar otras emociones (Figura 5.17b). Se pueden identificar varias burbujas de tristeza y sorpresa, pero también de alegría y anticipación. Hemos visto antes que el miedo es más frecuente en el sector de la salud que en otros sectores. Se pueden obtener conocimientos adicionales analizando la evolución de las emociones a lo largo del tiempo. En la Figura 5.17c podemos identificar que el miedo es particularmente alto durante octubre. Incluso podemos comparar diferentes emociones durante un período de interés. Por ejemplo, la Figura 5.17d muestra la presencia de miedo e ira durante el mes de octubre del 2017, con picos de miedo cerca del 23 de octubre del 2017. La ira también está presente cuando hay miedo, lo que puede implicar la correlación de estas dos emociones.

5.7.6. Análisis combinado

Algunos diagramas se utilizan para analizar diferentes aspectos de forma aislada, como en las secciones anteriores. Por ejemplo, los gráficos de líneas de la Sección 5.7.1 brindan una visión inmediata de la relación entre el nivel de actividad y el nivel de respuestas ciudadanas. En esas visualizaciones también podemos analizar el crecimiento de diferentes emociones presentes en los comentarios, para observar los estados de ánimo de los ciudadanos antes de que ciertos eventos sean desencadenados por los tweets del gobierno.

También podría ser de utilidad explorar la relación entre el enfoque de presentación de información adoptado por el gobierno y las respuestas ciudadanas. Para ello, la herramienta provee diagramas de Sankey para mostrar el flujo de recursos multimedia compartidos por las cuentas de Twitter de las secretarías y las correspondientes respuestas de los ciudadanos. La Figura 5.18 muestra la proporción de tweets publicados por una secretaria (línea vertical negra a la izquierda) que comparten una imagen (línea vertical en celeste),



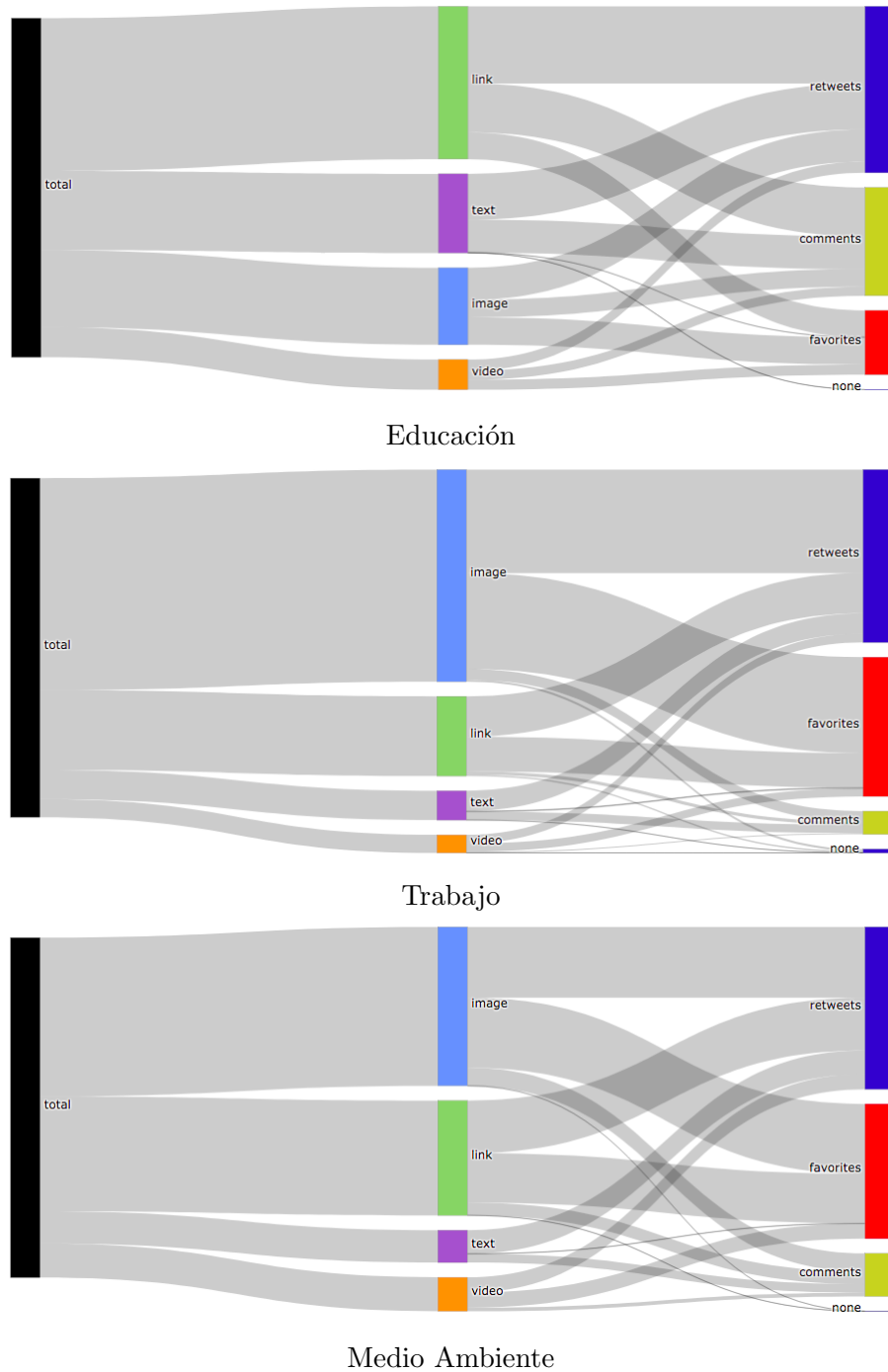


Figura 5.18: Relaciones entre los enlaces, vídeos, imágenes y texto compartidos por las cinco secretarías y los comentarios, *retweets* y favoritos de los ciudadanos.

un enlace (línea vertical en verde), un vídeo (línea vertical en naranja) o solo texto (línea vertical en violeta). Luego, para cada uno de ellos, podemos observar las respuestas de

los ciudadanos: *retweets* (en azul), favoritos (en rojo) o comentarios (en amarillo).

Por ejemplo, en la Figura 5.18, podemos observar que las Secretarías de Desarrollo Social, Salud y Trabajo comparten imágenes mientras que la Secretaría de Educación comparte en su mayoría enlaces. También podemos concluir que compartir vídeos no es una práctica muy común. Como hemos observado antes, las respuestas más comunes de los ciudadanos son los *retweets* y los tweets marcados como favoritos. La excepción se presenta en la Secretaría de Educación, donde la mayoría de los ciudadanos comentan los tweets publicados en lugar de marcarlos como favoritos. En general, se tiene un gran flujo de tweets que comparte imágenes, un flujo más pequeño que comparte enlaces y el resto comparte vídeos o solo texto. Si se observa el flujo entrante en las respuestas ciudadanas, entonces se conservan las proporciones. Por ejemplo, para la Secretaría de Desarrollo Social, podemos observar que una gran cantidad de *retweets* son en respuesta a los tweets con imágenes. Esto probablemente se deba a que el número de tweets con imágenes es mayor que el resto de tweets.

A través de la Figura 5.18 también podemos observar que algunos tweets no reciben ningún tipo de reacción o interacción de los ciudadanos, particularmente en el sector de trabajo. Recibir una gran cantidad de comentarios es deseable para todas las secretarías, ya que les permite conocer las opiniones de los ciudadanos sobre las políticas públicas que promueven. La secretaría con mejores resultados en este sentido es la Secretaría de Educación. Si bien esto puede deberse a un mayor interés de los ciudadanos en el sector de educación, las otras secretarías pueden analizar y adaptar la estrategia de la Secretaría de Educación para mejorar su interacción con la ciudadanía.

5.8. Discusión

En este capítulo se presentó una herramienta que facilita la recuperación de tweets y permite el análisis de las interacciones entre el gobierno y los ciudadanos utilizando diversas técnicas de visualización y análisis de sentimiento. El objetivo era proponer una metodología para este tipo de análisis y validarla examinando la actividad gubernamental y la respuesta ciudadana en el contexto de Twitter. La herramienta comprende todo un proceso que comienza con recopilar información de Twitter, procesar esta información y finalmente presentarla de manera amigable para facilitar su comprensión e interpretación. La herramienta se validó presentando dos casos de estudio: uno con cinco países de

Latinoamérica (ver Sección 5.6) utilizando la primera versión de la herramienta, y otro enfocándose en México (ver Sección 5.7) utilizando la última versión de la herramienta, que incorpora nuevas visualizaciones.

Las funcionalidades de la herramienta incluyen visualizaciones que permiten analizar los niveles de actividad del gobierno a lo largo del tiempo e identificar picos anormalmente altos o bajos, examinar las palabras usadas con mayor frecuencia y los temas discutidos más destacados, visualizar la estrategia de comunicación adoptada, y explorar las interacciones y emociones de la ciudadanía. Las técnicas implementadas también se pueden combinar, haciendo posible el análisis de características adicionales, como, por ejemplo, las relaciones entre los tipos de recursos multimedia compartidos y las respuestas de los ciudadanos. Dicho análisis combinado aumenta las capacidades de la herramienta para obtener conocimientos sobre cómo los ciudadanos perciben las decisiones gubernamentales o las políticas públicas, e identificar correlaciones entre las respuestas ciudadanas en términos de la polaridad e intensidad de las emociones expresadas y los anuncios gubernamentales, permitiendo potencialmente detectar qué temas generan reacciones polarizadas de los ciudadanos, entre otros.

A pesar de que las técnicas propuestas ofrecen un enfoque valioso para analizar la presencia del gobierno y la participación ciudadana en las redes sociales, la herramienta actualmente tiene algunas limitaciones. En la etapa de análisis, aplicamos un análisis de sentimiento simple que tiene algunas fallas. En particular, la conformación del lexicón puede influir de manera negativa. Si la cantidad de palabras asociadas con una emoción difiere significativamente de otras, entonces podríamos recibir resultados desequilibrados, como tener una gran cantidad de tweets relacionados con la «confianza» en cada análisis. También se necesita una nueva funcionalidad para detectar errores ortográficos, palabras utilizadas para expresar emociones opuestas, variaciones de palabras utilizadas localmente, y ambigüedades. Las nuevas versiones de la herramienta también deberían detectar *trolling* y *spamming* para limitar la influencia del ruido en el análisis.

Finalmente, se considera importante la recopilación de información adicional para ayudar a comprender el contexto de los valores mostrados, incluido el porcentaje de usuarios y las interacciones que se presentan en las visualizaciones en relación con la población total. Sin esto, la visualización puede conducir a una subestimación o sobreestimación de los valores mostrados, y la consiguiente mala interpretación de los resultados.

La herramienta permite identificar cómo los ciudadanos se comunican con las cuentas del gobierno: cuánto les gusta el contenido publicado (favoritos), cómo analizan las publicaciones (*retweets*) y qué piensan acerca de la información publicada (comentarios). Para futuros trabajos sería interesante agregar funcionalidades para analizar los picos en los comentarios recuperados, y realizar un análisis emocional de los *retweets*, considerando que estos pueden ser considerados positivos o negativos dependiendo del texto que los acompañe. Otras funcionalidades que pueden mejorar la herramienta están relacionadas a la exploración de otras técnicas de visualizaciones y de análisis combinados, como puede ser el análisis de sentimientos para identificar correlaciones entre tipos de emociones en las respuestas a anuncios de políticas particulares, o una funcionalidad para detectar los temas que generaron las reacciones más polarizadas de los ciudadanos.

Otras funcionalidades que podrían enriquecer la herramienta en un trabajo futuro es la recuperación de tweets relacionados a eventos de interés del momento mediante el uso de hashtags o palabras clave, explorar otros algoritmos de procesamiento del lenguaje natural para descubrir los principales temas discutidos, y mejorar el léxico utilizado para el análisis de sentimientos mediante la adaptación a dominios de estudio específicos. El análisis de sentimiento puede mejorarse si se refina el concepto de emociones para una mejor comprensión de su semántica en cada sector gubernamental; por ejemplo, definir el significado de «confianza» en los sectores ambiental y laboral. Finalmente, las técnicas presentadas se pueden aplicar a otras plataformas de redes sociales como Facebook o Instagram, considerando sus propios mecanismos de interacción entre los usuarios y los métodos de recuperación disponibles a través de las API respectivas.

Capítulo 6

Herramientas informáticas para análisis de comunicaciones públicas

En este capítulo se presenta una plataforma para recuperar información y refinar los resultados relevantes para un usuario experto mediante la aplicación de aprendizaje activo. La Sección 6.1 presenta la motivación. La Sección 6.2 explica la metodología de desarrollo, y los requerimientos y las funcionalidades de la herramienta se abordan en la Sección 6.3. En las secciones 6.4 y 6.5 se indican el diseño, la implementación y el testeo de la plataforma. La validación de la herramienta se realiza mediante dos casos de estudio vinculados a la recuperación de noticias relevantes respecto a la violencia de género y la criminalización del trabajo sexual, detallados en las secciones 6.6 y 6.7 respectivamente. Finalmente, las contribuciones y las limitaciones de la plataforma desarrollada se discuten en la Sección 6.8.

El trabajo descrito en este capítulo se realizó en colaboración directa con el Dr. Evangelos Milios de Dalhousie University y la Dra. Evangelina Tastsoglou de Saint Mary's University (Halifax, Canadá) durante una estancia de cinco meses de la tesista en Dalhousie University gracias a la beca Emerging Leaders in the Americas Program.

6.1. Motivación

Como se ha indicado en capítulos anteriores, los medios digitales, y en particular las redes sociales, se han convertido en una herramienta muy utilizada tanto por gobiernos

como por ciudadanos para compartir noticias e información en general. La recuperación y análisis de esta información comienza a cobrar relevancia para la toma de decisiones de gobierno. Sin embargo, la diversidad de fuentes y de información disponible dificulta la tarea de centralizar y recopilar aquellos ítems relevantes para un análisis posterior.

Este estudio está impulsado por la pregunta de investigación definida previamente en la Sección 1.2: **¿Qué componentes y técnicas pueden ser parte de una infraestructura de software que haga uso de información volcada en plataformas digitales para facilitar la toma de decisiones por parte de funcionarios públicos?** Se busca el diseño e implementación de una herramienta que incluya un método para que los usuarios puedan recabar información de diferentes fuentes y quedarse con el subconjunto que es relevante para el análisis que busca llevar a cabo.

En las secciones posteriores, se presenta una herramienta de software para recuperar información de diversas fuentes y obtener un conjunto de noticias relevantes utilizando un modelo de clasificación. Esta herramienta funciona como una herramienta complementaria encargada del primer paso en el análisis de información: la recuperación y la identificación de noticias relevantes. Las técnicas utilizadas y las diversas funcionalidades de la herramienta se presentarán en secciones posteriores.

6.2. Metodología de desarrollo

El desarrollo de la herramienta siguió una metodología de desarrollo evolutiva. Las principales características de la metodología elegida y la justificación de su elección se explicó en la Sección 4.2. A diferencia de las herramientas anteriores, donde los requerimientos fueron definidos de forma exploratoria por la tesista y el equipo de investigación, los requerimientos en este caso se obtuvieron de usuarios reales. Se definieron las funcionalidades de la plataforma en base a estos requerimientos, y fueron implementados y mejorados según la retroalimentación del usuario en los sucesivos ciclos. En las secciones posteriores se presentan las características más importantes del desarrollo.

6.3. Requerimientos

El diagrama de caso de uso de la Figura 6.1 muestra las funcionalidades de la herramienta: *Crear modelos*, *Buscar información*, *Clasificar ítems* y *Exportar ítems*. Las funcionalidades de *Visualizar ítems del modelo* y *Actualizar información del ítem* extienden la funcionalidad *Clasificar ítems*. La mayoría de las funcionalidades implican la funcionalidad de *Seleccionar modelo* y de *Seleccionar usuario*. El detalle de todas las funcionalidades se abordará en las siguientes subsecciones.

La herramienta puede ser utilizada por cualquier usuario experto en algún dominio. Por ejemplo, un sociólogo que buscará noticias relacionadas a su trabajo de investigación, o un oficial de gobierno que busca evaluar el impacto de una política pública buscando noticias y comentarios de ciudadanos en diferentes fuentes de información. El usuario puede crear modelos de clasificación relacionados a cierta temática, realizar búsquedas en diferentes fuentes de información mediante el uso de palabras clave, y refinar las predicciones de los modelos creados a fin de obtener solo las noticias relevantes para su posterior análisis.

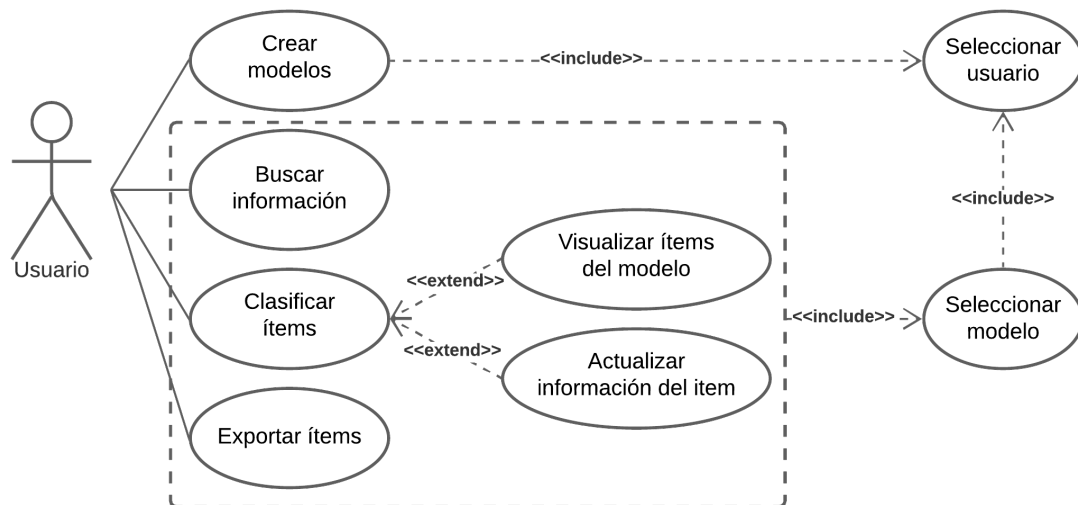


Figura 6.1: Diagrama de caso de uso.

6.3.1. Seleccionar usuario

Antes de comenzar a utilizar las diferentes funcionalidades de la herramienta, debe ingresarse un nombre de usuario, y se constatará que exista en la base de datos antes de continuar con el uso de la herramienta. Cada usuario tendrá asociado una serie de modelos

de clasificación. El usuario actual puede cambiarse en cualquier momento. Actualmente, la herramienta no permite la creación de nuevos usuarios. Esto solo puede hacerlo el administrador de la herramienta.

6.3.2. Crear modelos

La herramienta permite al usuario crear modelos de clasificación. Cada modelo representará una «temática» y estará identificado por un nombre único, que se controlará con aquellos modelos existentes en la base de datos. Además de dar un nombre, el usuario debe proveer un conjunto pequeño de ejemplos relevantes e irrelevantes de URLs a noticias, posts de Reddit o tweets (por ejemplo, cuatro ejemplos relevantes y cuatro no relevantes). Estos ejemplos se utilizarán para entrenar el modelo por primera vez. Una vez creado y entrenado el modelo, este será el seleccionado para trabajar.

6.3.3. Seleccionar modelo

El usuario de la herramienta puede elegir crear un modelo nuevo (utilizando la funcionalidad anteriormente descrita) o seleccionar uno de los ya creados y asociados al usuario actual, que se mostrarán mediante una lista desplegable. El modelo puede cambiarse en cualquier momento. Todas las búsquedas y las clasificaciones manuales a los ítems de información (mediante el ciclo de aprendizaje activo) que se realicen durante la sesión se asociarán al modelo actualmente elegido.

6.3.4. Buscar información

La herramienta permite realizar búsquedas de dos formas: mediante el uso de palabras clave o por medio de URLs provistas por el usuario. Para la primera opción, debe seleccionarse dónde se desea buscar ítems de información (Google, Reddit o Twitter) y proveer las palabras clave para la búsqueda. Para la segunda opción, el usuario debe ingresar una serie de URLs separados por espacios.

La herramienta realizará las búsquedas correspondientes, recuperando la información pertinente de los diferentes ítems de noticias. Cada uno de estas noticias, posts o tweets se asociará al modelo actualmente elegido por el usuario.

6.3.5. Clasificar ítems

A fin de obtener los ítems de información más relevantes, el usuario deberá clasificar manualmente un número reducido de noticias, posts o tweets. La herramienta mostrará los ítems de información que el modelo de clasificación no sabe cómo clasificar, es decir los ítems cuya probabilidad de ser relevante es aproximadamente del 50%. El usuario puede cambiar este rango. Por ejemplo, puede elegir ver aquellas noticias que el modelo considera poco relevantes (20-30% de probabilidad de relevancia) o muy relevantes (80-90%).

Del conjunto de noticias que la herramienta recupera dentro del rango elegido, el usuario clasificará manualmente cuatro de ellas. Esto significa que verá el ítem de información y decidirá, según su conocimiento, si es o no relevante para su trabajo. Una vez clasificados los ítems, la herramienta volverá a entrenar el modelo con esta nueva información. Esto puede hacerse cuantas veces el usuario considere necesario. Esta tarea constituye el ciclo de aprendizaje activo.

6.3.6. Visualizar ítems del modelo

La herramienta provee una opción de visualizar el comportamiento del modelo en un diagrama de dispersión, como se ilustra en la Figura 6.2. En este diagrama, los puntos azules son los ítems de información (noticias, posts de Reddit o tweets) no clasificados por el usuario, y los puntos naranjas son los ítems ya clasificados por el usuario. El eje x representa la relevancia del ítem según el modelo de clasificación: de no relevante (0%) a relevante (100%). El eje y representa la cantidad de palabras del ítem de información. En la leyenda del gráfico se indican la cantidad de ítems clasificados por el usuario (48) y por el modelo (4569).

Este gráfico permite identificar el comportamiento del modelo a medida que se va entrenando en el ciclo de aprendizaje activo. Un buen indicativo de que se requiere seguir entrenando el modelo es si la mayoría de los ítems de información están cerca del 50% de relevancia. Por otro lado, si el gráfico se ve muy concentrado en los extremos, con una pequeña cantidad de ítems en el medio, se confirmaría el éxito del modelo y no sería necesario seguir entrenándolo.

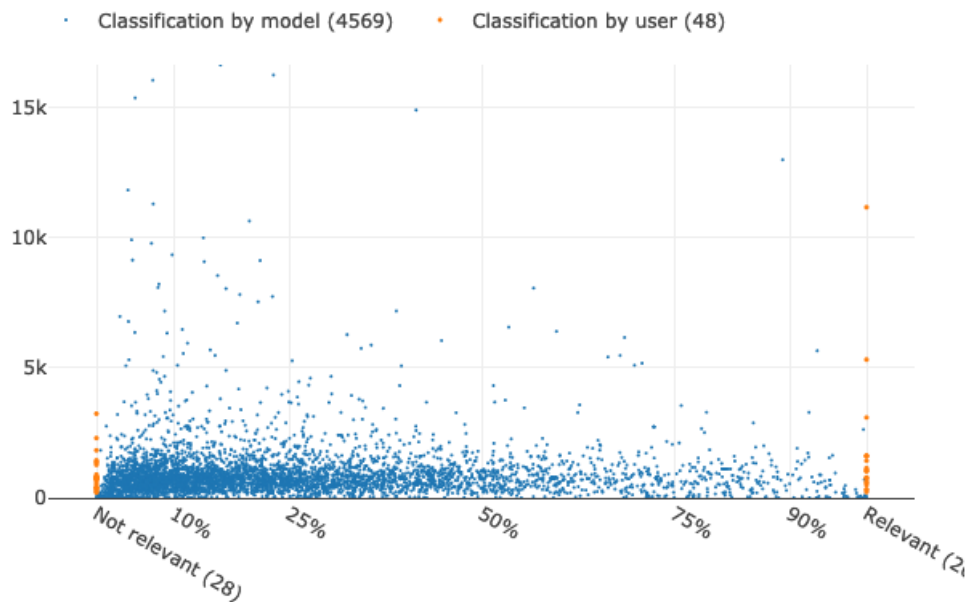


Figura 6.2: Diagrama de dispersión que indica el comportamiento del modelo.

6.3.7. Actualizar información del ítem

Cuando el usuario está evaluando si una noticia, post o tweet es relevante para su trabajo, tiene la opción de actualizar la información de ese ítem (como puede ser título o contenido), en el caso de que la herramienta haya recuperado datos erróneos o incompletos.

6.3.8. Exportar ítems

La herramienta permite exportar los ítems de información de dos formas diferentes. La primera es la creación de un archivo ZIP que contendrá archivos TXT. Cada archivo será un ítem de información diferente dentro de un rango de relevancia elegido por el usuario. De esta forma, el usuario puede elegir exportar aquellos ítems cuya relevancia sea mayor a un determinado valor (por ejemplo, mayor a 80%).

La segunda opción es exportar ítems hacia un nuevo modelo de clasificación, eligiendo un nombre para el nuevo modelo y un rango de relevancia para seleccionar los ítems a exportar. Esto es especialmente útil si el usuario desea refinar su búsqueda de ítems relevantes sin «confundir» al modelo al darle clasificaciones contradictorias o ambiguas. De este modo, el usuario puede tener múltiples modelos que pueden llegar a representar diferentes ramas dentro de una misma temática general.

6.4. Diseño

Esta sección presenta las principales decisiones de diseño del sistema: la Sección 6.4.1) indica la arquitectura del sistema, y la Sección 6.4.2 presenta el modelo de datos.

6.4.1. Diseño de la aplicación

Tal como en las herramientas anteriores, el diseño de la estructura se basa en un estilo cliente-servidor. El motivo por el cual fue elegida esta arquitectura se indica en la Sección 4.4.1. La Figura 6.3 muestra el diagrama de componentes de la herramienta. Aquellos componentes presentes en el paquete *Frontend* se ejecutan en el lado del cliente, mientras que el resto se ejecuta en el servidor.

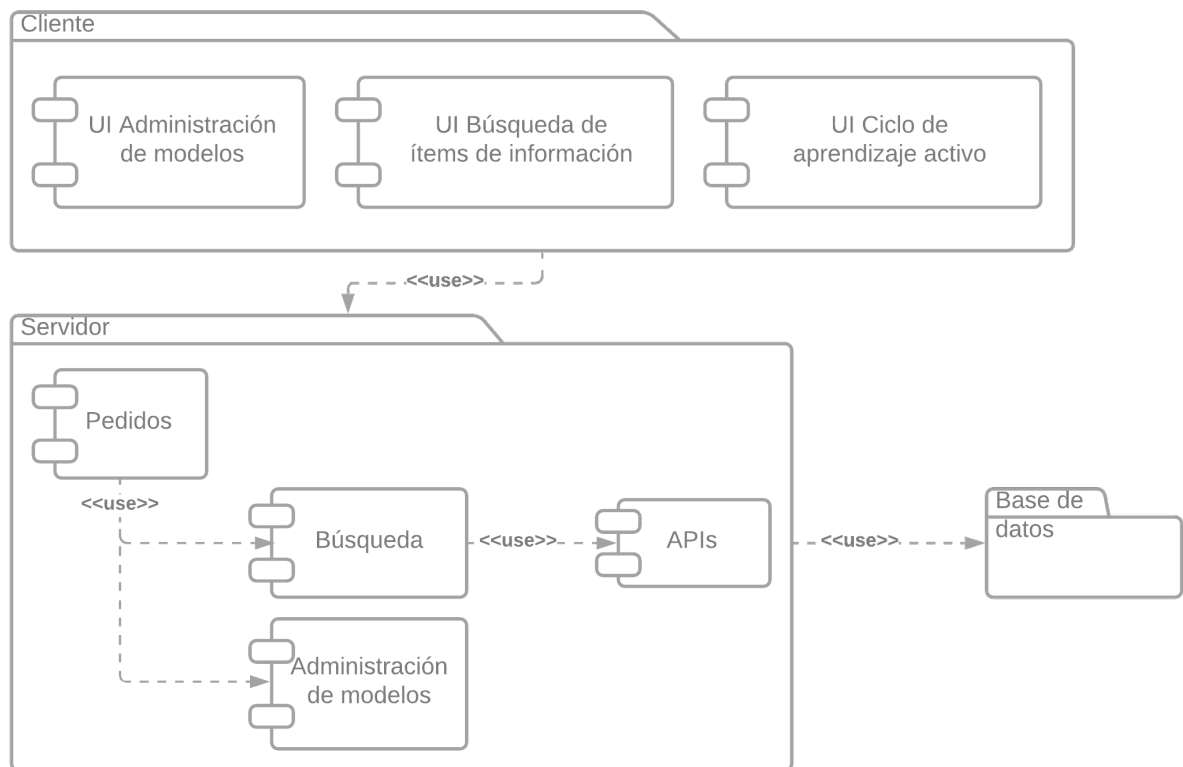


Figura 6.3: Diagrama de componentes.

El diseño del sistema tiene tres paquetes principales: *Cliente*, *Servidor* y *Base de datos*. El *Cliente* provee los componentes que implementan todas las funcionalidades del sistema

que pueden accederse mediante el navegador. Los componentes principales de la interfaz de usuario son *UI Administración de modelos* para crear y seleccionar los modelos que utilizará el usuario, *UI Búsqueda de ítems de información* para buscar noticias, posts de Reddit o tweets, y *IU Ciclo de aprendizaje activo* para clasificar los ítems de información y visualizar el comportamiento del modelo.

Por otro lado, el paquete *Servidor* implementa el servidor del sistema. Sus componentes principales son: (1) *Pedidos*, por donde se reciben y resuelven los pedidos del cliente, devolviendo la información esperada, (2) *Búsqueda*, que realizará las búsquedas utilizando las APIs correspondientes y guardará la información de forma adecuada, y (3) *Administración de modelos*, que administrará la creación de modelos y se encargará del entrenamiento de los mismos ante las actualizaciones del usuario. El componente *APIs* se ocupa de las conexiones con las APIs de Twitter, Reddit y la conexión con Google Noticias. Finalmente, el paquete *Base de datos* se encarga del manejo de la base de datos, incluidas las operaciones de guardar, obtener y borrar elementos.

6.4.2. Modelo de datos

La herramienta desarrollada es capaz de recopilar datos de Google Noticias, Twitter y Reddit por medio de librerías que realizan conexiones con las respectivas APIs. Las diferentes librerías utilizadas permiten recuperar noticias, tweets y posts utilizando palabras clave o URLs, junto con datos adicionales a estos ítems de información, como pueden ser contenido, título, fecha de creación, y, en el caso de posts y tweets, información del usuario, *likes/upvotes* y comentarios, entre otros.

La base de datos tiene tres colecciones diferentes para guardar datos de cada noticia, tweet o post de Reddit. En el caso de la colección de noticias, se guardan para cada ítem los siguientes datos:

- *url*: URL de la noticia.
- *creation_time*: el día y horario de creación.
- *title*: el título de la noticia.
- *text*: el contenido de la noticia.

Para cada tweet se guardan los siguientes datos:

- *url*: URL del tweet.

- *tweet_id*: número identificador del tweet.
- *user*: información acerca del usuario. En particular:
 - *user_id*: número identificador del usuario.
 - *username*: nombre de usuario.
- *text*: el contenido del tweet.
- *creation_time*: el día y horario de creación del tweet.
- *hashtags*: una lista de los hashtags mencionados en el tweet.
- *user_mentions*: una lista de los usuarios mencionados en el tweet, donde cada elemento de la lista contiene el identificador y el nombre del usuario.
- *is_retweeted*: un campo *flag* para indicar si el tweet es un *retweet* o no. Sus valores pueden ser *verdadero* o *falso*.
- *favorites*: el número de favoritos del tweet.
- *retweets*: el número de *retweets* del tweet.
- *comments*: una lista con los comentarios realizados. Cada comentario es un tweet, por lo tanto tiene los mismos datos aquí listados.

Finalmente, para cada post de Reddit se guardarán los siguientes datos:

- *url*: URL del post.
- *user*: nombre de usuario.
- *title*: título del post.
- *text*: el contenido del post.
- *creation_time*: el día y horario de creación del tweet.
- *upvotes*: el número de votos positivos del post.
- *comments*: una lista con los comentarios realizados. Cada comentario es un post en sí mismo, por lo tanto tiene los mismos datos aquí listados.

Además, para cada ítem de información se guarda también:

- *search_keywords*: una lista de las palabras clave utilizadas para recuperar la noticia, tweet o post. Un mismo ítem puede aparecer como resultado de diferentes búsquedas con diferentes palabras clave.
- *models*: una lista de los modelos que utilizan este ítem junto a la clasificación definida por el usuario. Esta clasificación puede ser *relevante*, *no relevante* o *nula*, si el usuario no clasificó este ítem. Una misma noticia, tweet o post puede ser utilizada por diferentes modelos.

- *datatype*: el tipo de fuente de información. Puede ser *google*, *twitter* o *reddit*.

Así mismo, existen otras dos colecciones: una para los modelos y otra para los usuarios.

Para cada usuario se guarda la siguiente información:

- *username*: el nombre de usuario.
- *models*: una lista de los modelos que creados y utilizados por el usuario.

La colección de los modelos guardan la siguiente información:

- *original_name*: el nombre del modelo dado por el usuario.
- *model_name*: el nombre del modelo modificado por el servidor (en minúsculas y cambiando los espacios por guiones bajos).
- *model_file*: la dirección donde se guarda el modelo.

El detalle de los datos utilizados se indicarán en las Secciones 6.6 y 6.7.

6.5. Implementación y testeo

Para el desarrollo del *frontend* se utilizó el framework Angular¹ (9.1.0), además de Angular Material² (9.2.4) como librería de estilo. Para las visualizaciones se usó la librería *angular-plotly*³ (2.0.0).

Por otro lado, el *backend* se implementó en Python⁴ (3.8.1), usando *flask*⁵ (1.1.2) para la creación de la aplicación web y *rq*⁶ (1.4.2) para la administración de tareas en *background*. Las librerías utilizadas para la recuperación de información de Google Noticias fueron *googlesearch*⁷ (2.0.3) y *newspaper3k*⁸ (0.2.8); para Reddit se utilizó *praw*⁹ (7.0.0), y para Twitter se usaron *tweepy*¹⁰ (3.8.0) y *twarc*¹¹ (1.8.3). Se utilizó la librería *pandas*¹²

¹ Angular - <https://angular.io/>
² Angular Material - <https://material.angular.io/>
³ angular-plotly - <https://github.com/plotly/angular-plotly.js>
⁴ Python - <https://www.python.org/>
⁵ flask - <https://flask.palletsprojects.com/en/1.1.x/>
⁶ rq - <https://github.com/rq/rq>
⁷ googlesearch - <https://github.com/MarioVilas/googlesearch>
⁸ newspaper3k - <https://github.com/codelucas/newspaper>
⁹ praw - <https://praw.readthedocs.io/en/latest/>
¹⁰ tweepy - <https://www.tweepy.org/>
¹¹ twarc - <https://github.com/DocNow/twarc>
¹² pandas - <https://pandas.pydata.org/>

(1.0.3) para la manipulación de los datos y `fasttext`¹³ (0.9.2) para el algoritmo de clasificación. Además, se empleó `mongoDB`¹⁴ como bases de datos, haciendo uso de la librería `pymongo`¹⁵ (3.10.1) para establecer la conexión.

El testeo de la herramienta se realizó a nivel de pruebas unitarias llevadas a cabo por la tesista. Las pruebas de integración y de uso fueron verificadas por las directoras. Además, el testeo se complementó con demos realizadas a partes interesadas y usuarios de la plataforma. Finalmente, la validación de la herramienta se llevó a cabo mediante dos casos de estudio, que se describen en las siguientes secciones.

6.6. Caso de estudio: Noticias de Google

La primera validación de la herramienta se realizó como parte del desarrollo de una tesis de magister en Desarrollo Internacional en Saint Mary's University (Halifax, Canadá), bajo la supervisión de la Dra. Evangelina Tastsoglou. El usuario experto es un Bachiller Universitario en Ciencias en Relaciones Internacionales. Su objetivo consistía en recopilar noticias relevantes para el desarrollo de su tesis, en particular noticias relacionadas a experiencias de violencia doméstica de mujeres migrantes en Canadá.

El primer paso en esta validación consistió en la definición del tema de investigación por parte del usuario experto del dominio: establecer el tipo de noticias buscadas y realizar un trabajo exploratorio en Google para determinar qué palabras clave podrían ser útiles para la búsqueda dentro de la herramienta desarrollada. Este trabajo exploratorio consistía en el uso de diferentes palabras clave y el análisis de unos pocos artículos de Google Noticias. Una vez realizado este trabajo, el usuario creó un modelo utilizando la herramienta: se le asignó un nombre (*domestic_violence*) y se realizó un primer entrenamiento con tres ejemplos de noticias relevantes y tres ejemplos irrelevantes. Estas noticias fueron obtenidas de la búsqueda exploratoria realizada por el usuario.

A continuación, se realizó una búsqueda en Google Noticias mediante la herramienta utilizando «*domestic violence*» como palabras clave, obteniendo 289 resultados. Posteriormente, el usuario comenzó con el ciclo de entrenamiento activo: la herramienta le otorgó cuatro noticias para que clasifique manualmente. Mediante este proceso, el usuario notó

¹³ `fasttext` - <https://fasttext.cc/>

¹⁴ `mongoDB` - <https://www.mongodb.com/es>

¹⁵ `pymongo` - <https://pymongo.readthedocs.io/>

que todas las noticias devueltas eran de Estados Unidos o del Reino Unido, cuando se deseaban noticias de Canadá. Por lo tanto, se realizó otra búsqueda utilizando las palabras clave «*domestic violence canada*», resultando en 264 noticias nuevas. Se prosiguió luego con la clasificación manual por parte del usuario durante tres ciclos.

De este primer uso de la herramienta, se determinó que la plataforma trabajaba rápidamente para clasificar y re-entrenar el modelo, pero que tardaba un tiempo considerable para la búsqueda de noticias debido a las limitaciones de la API. También por limitaciones ajenas a la herramienta desarrollada, solo se logró recuperar una pequeña porción de todas las noticias conteniendo las palabras clave (siempre en torno a los 250-300 resultados). Para solventar este problema se decidió realizar múltiples búsquedas utilizando las mismas palabras clave («*domestic violence*») y agregándole diferentes provincias de Canadá (dos ejemplos de búsquedas son «*domestic violence alberta*» y «*domestic violence nova scotia*»). El resultado final fue de 2953 noticias acerca de violencia doméstica en Canadá. Luego, se volvió al ciclo de aprendizaje activo para que el usuario pueda clasificar manualmente más noticias.

Dado que no se observaban grandes cambios en el comportamiento del modelo a pesar de los ciclos de clasificación manual, se optó por modificar los parámetros del clasificador, logrando un mayor aprendizaje con menos esfuerzo. Aparte, notamos que el usuario había clasificado erróneamente algunas noticias dado que se basó en el *lugar* de la noticia en vez del *contenido* (es decir, clasificó como irrelevante a noticias relevantes solo porque no pertenecían a Canadá). Sumado a esto que el enfoque del usuario respecto al tipo de noticias deseadas sufrió algunos cambios en el proceso, se volvieron a revisar aquellas noticias clasificadas para asegurarnos de darle al modelo la información correcta.

Una vez terminadas las tareas anteriores, se observaron cambios positivos en el modelo y en la cantidad de noticias relevantes, de modo que se tomó la decisión de finalizar con la clasificación manual por parte del usuario. La evolución en el comportamiento del modelo puede verse en la Figura 6.4, donde se puede observar cómo se comporta cuando no hay clasificación por parte del usuario, y luego de 1, 2, 3, 6 y 13 ciclos de clasificación manual. En estos histogramas, el eje x representa el porcentaje de relevancia de las noticias, y el eje y el número de noticias. Sin clasificación por parte del usuario, se observa que la mayoría de las noticias están concentradas en el medio, es decir en el 50 % de relevancia. A medida que pasan los ciclos de aprendizaje activo, con clasificación manual por parte del usuario, el porcentaje de relevancia dado por el modelo va moviéndose hacia los extremos. Se puede

observar que una gran cantidad de noticias son poco relevantes. También es evidente que no hay una gran diferencia luego de tres ciclos de clasificación manual, aunque luego de 13 ciclos se nota una considerable concentración de noticias en el extremo más relevante (95-100% de relevancia).

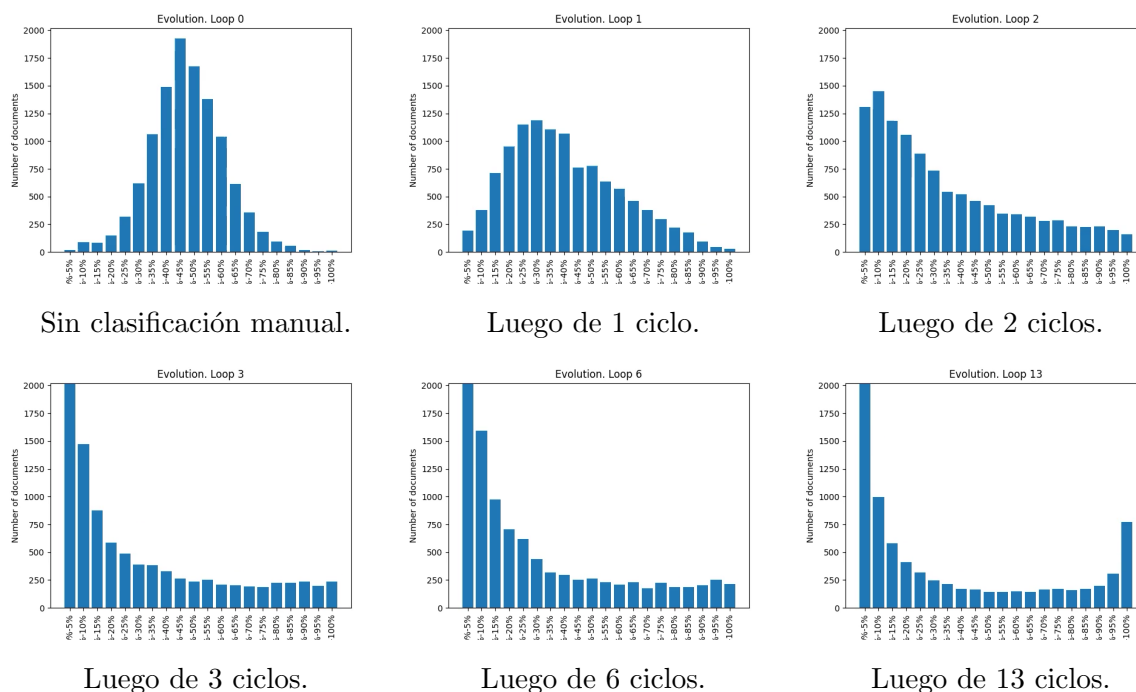


Figura 6.4: Evolución del comportamiento del modelo *domestic_violence*.

El usuario exportó las noticias (como documentos TXT) con relevancia mayor al 99,5% para analizarlas utilizando una herramienta de visualización de agrupaciones de documentos por temáticas. El número de noticias resultantes fue de 737 (Grupo A). De estas, y realizando un análisis más detallado del contenido, el usuario recuperó 104 útiles para su trabajo de investigación.

Cabe destacar que en su momento el usuario consideró volver al proceso de aprendizaje activo creando un nuevo modelo para mejorar el resultado final. Para esto, se utilizaron noticias del primer modelo: se exportaron las noticias con relevancia mayor al 85% del modelo *domestic_violence*, se asignó un nombre al nuevo modelo (*domestic_violence_migrant_women*), y el usuario seleccionó cinco ejemplos relevantes y cinco irrelevantes para entrenar este modelo (extraídas del Grupo A), y comenzó con los ciclos de clasificación manual. El usuario también consideró recuperar posts de Reddit utilizando palabras clave como «*forced marriage*», «*my story on abuse*» o «*sexual abuse*», entre

otros, para complementar las noticias de este segundo modelo. Sin embargo, finalmente utilizó aquellas noticias del Grupo A, haciendo uso de la herramienta de visualización de agrupaciones temática para aliviar la carga de selección del conjunto final de noticias.

6.7. Caso de estudio: Noticias de Google

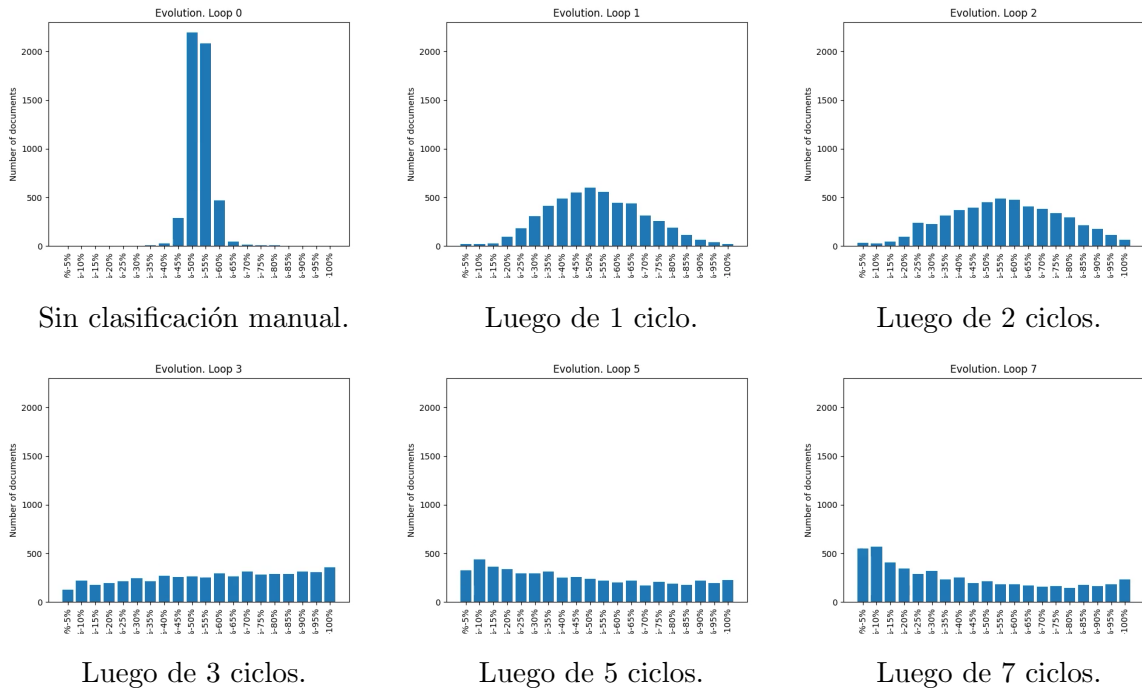
La segunda validación de la herramienta se realizó como parte del desarrollo de una tesis de magister en Desarrollo Internacional en Saint Mary's University (Halifax, Canadá), bajo la supervisión de la Dra. Evangelina Tastsoglou. El usuario experto, un Bachiller Universitario en Letras en Estudios Internacionales. Su objetivo era recuperar noticias relevantes al desarrollo de su tesis, específicamente aquellas relacionadas al impacto en las trabajadoras sexuales luego de la criminalización del trabajo sexual en Cambodia.

Al igual que en el caso de estudio anterior, la primera tarea consistió en la definición de la temática, estableciendo con exactitud el tipo de noticias buscadas, y en la realización de un trabajo exploratorio en fuentes de información (como Google Noticias) para determinar las palabras clave que podrían dar resultados satisfactorios para la búsqueda mediante la herramienta desarrollada. Una vez completo ese trabajo, el usuario creó un modelo de clasificación en la herramienta: le asignó un nombre (*criminalization_sex_work*) y diez ejemplos de noticias relevantes y diez ejemplos irrelevantes. Estas noticias fueron obtenidas de la búsqueda exploratoria realizada por el usuario.

A continuación, se comenzó con las búsquedas en Google Noticias mediante la herramienta. Se utilizaron las palabras clave «*sex workers cambodia*», recuperando 463 noticias. Teniendo en cuenta la experiencia del caso de estudio anterior, se realizaron también búsquedas con cada provincia de Cambodia (por ejemplo, «*sex workers battambang*»), obteniendo un total de 3896 noticias. Se prosiguió luego al ciclo de aprendizaje activo. Considerando que no habían suficiente noticias relevantes en Cambodia, el usuario decidió realizar búsquedas en Tailandia («*sex workers thailand*»), incluyendo también las ciudades más importantes. El total de noticias fue de 5144.

El usuario realizó un total de siete ciclos de clasificación manual antes de concluir que era suficiente. La Figura 6.5 muestra el comportamiento del modelo cuando no hay clasificación por parte del usuario, y luego de 1, 2, 3, 5 y 7 ciclos de clasificación manual. Al principio, todas las noticias se encuentra en el medio del histograma (50 % de relevancia),

y, a medida que se clasifican noticias manualmente, empiezan a alejarse del medio, hasta que casi no se observa cambios de un ciclo al siguiente.



Sin clasificación manual.

Luego de 1 ciclo.

Luego de 2 ciclos.

Luego de 3 ciclos.

Luego de 5 ciclos.

Luego de 7 ciclos.

Figura 6.5: Evolución del comportamiento del modelo *criminalization_sex_work*.

Para la fase exploratoria utilizando la herramienta de visualización de agrupaciones por temática, el usuario decidió exportar todas las noticias del modelo con relevancia mayor al 80%, siendo este un total de 625. Luego de explorar las noticias, consideró que sería mejor volver a limitar el alcance a las noticias pertenecientes solo a Cambodia. De este modo, redujo la cantidad de noticias a 488. Después del análisis detallado, y utilizando la herramienta de visualización de agrupaciones por temática, el usuario obtuvo 120 noticias relevantes para su investigación.

6.8. Discusión

En este capítulo se presentó una herramienta que permite (1) la recuperación de información de diferentes fuentes, en particular de Google Noticias, Reddit y Twitter, y (2) el refinamiento de resultados relevantes por medio de un sistema de aprendizaje activo. El objetivo era proponer una metodología a fin de facilitar el trabajo de usuarios

expertos al recuperar información de diversas fuentes sobre temáticas específicas. La herramienta comprende la creación de un modelo de clasificación, la recuperación de ítems de información de diversas fuentes, y el entrenamiento del modelo de clasificación. Este entrenamiento requiere poco esfuerzo por parte del usuario: clasificando manualmente un pequeño número de noticias, el modelo de clasificación aprende y clasifica un gran número de noticias, permitiendo seleccionar solo las más relevantes para el usuario.

A fin de validar esta herramienta, se realizaron dos casos de estudio con usuarios expertos en las secciones 6.6 y 6.7. En estos casos, los usuarios fueron estudiantes de máster de Desarrollo Internacional, pero el usuario de la herramienta puede ser cualquier usuario experto en alguna temática, incluidos en este grupo cualquier funcionario público. Un oficial de gobierno podría buscar posts o tweets referidos a cierta política pública o problemática social a fin de analizar la opinión ciudadana, o incluso identificar las decisiones tomadas por otros gobiernos y su impacto en la sociedad. La herramienta puede jugar un rol complementario con el uso de otra herramienta, ya sea para identificar, analizar o visualizar temáticas y contenido.

La herramienta permite crear cualquier número de modelos de clasificación, de modo que el usuario pueda enfocarse en temáticas específicas. Además, tiene la opción de exportar las noticias relevantes (1) de un modelo a un modelo nuevo, ramificando la temática, o (2) como un ZIP, pudiendo utilizar los archivos en una herramienta complementaria o accediendo a su contenido directamente. El ciclo de aprendizaje activo, donde el usuario clasificará manualmente un número reducido de noticias, permite además observar el comportamiento del modelo a medida que se realizan ciclos de clasificación manual. Esto ayuda al usuario a determinar cuándo finalizar con la clasificación manual, y a tener una idea aproximada de la cantidad de noticias relevantes que obtendrá finalmente.

La herramienta tiene buenos tiempos de respuesta al clasificar noticias y visualizar el gráfico de comportamiento del modelo, es decir, la distribución de las noticias en el eje de relevancia. Sin embargo, por limitaciones de conexiones con las APIs, la búsqueda mediante palabras clave tarda un tiempo considerable. De todos modos, el usuario puede continuar con la clasificación manual mientras se realiza la búsqueda de noticias.

De la interacción con los usuarios de los casos de uso nacieron algunas funcionalidades extras en la herramienta, como la opción de exportar noticias hacia un nuevo modelo de clasificación. Otra funcionalidad que podría agregarse en un trabajo futuro, que surge también de la retroalimentación del usuario, es la posibilidad de eliminar las noticias que

se obtuvieron de determinadas búsquedas (es decir, noticias recuperadas utilizando ciertas palabras clave). También se considera agregar nuevas fuentes de información (como puede ser otras redes sociales) y permitir importar documentos.

Capítulo 7

Conclusiones y trabajo a futuro

En este capítulo se resume el trabajo desarrollado en base a las preguntas de investigación definidas previamente (Sección 7.1), y se presentan las contribuciones y limitaciones de las herramientas implementadas (Sección 7.2). Luego, se listan las publicaciones que surgieron como parte del desarrollo de esta tesis (Sección 7.4). Finalmente, se indica el trabajo a futuro que puede realizarse en base al contenido presentado (Sección 7.3).

7.1. Este trabajo

El trabajo presentado en esta tesis estuvo guiado por las preguntas de investigación definidas en la Sección 1.2. La PI1 —*¿Cuáles son las técnicas de minería de datos y de visualización más convenientes para el análisis de opiniones, quejas y reclamos vertidos en diversas plataformas digitales en el contexto de participación electrónica?*— se respondió realizando un estudio detallado del estado del arte, desarrollado en el Capítulo 3. Este estudio comenzó con la búsqueda de investigaciones relevantes para identificar y presentar diferentes aspectos de e-Participación, como son las dimensiones, las técnicas utilizadas, y los principales problemas y desafíos al momento de desarrollar iniciativas (ver Sección 3.1). Se continuó con la búsqueda de artículos científicos que presentan plataformas de e-Participación y su correspondiente análisis a fin de identificar las técnicas adecuadas para usar en herramientas en el contexto de participación ciudadana (ver Sección 3.2). De este análisis derivó una clasificación de iniciativas que tiene en cuenta su nivel y área de participación, que permite un posterior estudio de la relación de estos aspectos con las

técnicas utilizadas (ver Sección 3.5), y que indica que ciertas tecnologías pueden ser más útiles en ciertos niveles y áreas, y dependiendo del objetivo de la iniciativa.

Por otro lado, la PI2 —*¿Qué componentes y técnicas pueden ser parte de una infraestructura de software que haga uso de información volcada en plataformas digitales para facilitar la toma de decisiones por parte de funcionarios públicos?*— se respondió mediante el desarrollo de tres herramientas de software en el contexto de participación electrónica. El desarrollo de estas herramientas estuvo también ligado al trabajo realizado para responder la PI1, dado que uno de los pasos principales para el diseño e implementación de estos prototipos fue la selección de algoritmos de minería de datos y técnicas de visualización apropiados acordes al objetivo buscado.

La primera herramienta se presentó en el Capítulo 4, centrada en el uso de PQR ciudadanos de un municipio. Este tipo de datos permite la adopción de algoritmos que tengan en cuenta su componente espacial, y el uso de visualizaciones en mapas de la ciudad. En particular, los métodos de agrupamiento utilizados fueron DBSCAN y *hierarchical greedy clustering*, y se presentaron visualizaciones de mapa de calor y de agrupaciones. Estos métodos de agrupamiento son aptos para formar parte de una infraestructura de software a fin de facilitar la obtención de información relevante para el proceso de toma de decisiones del gobierno al momento del análisis y resolución de los PQR. Así mismo, se cumplieron los dos objetivos planteados. Primeramente se desarrolló el prototipo que implementa el sistema propuesto, cuyo diseño y tecnologías utilizadas se indican en las secciones 4.4 y 4.5. El segundo objetivo consistía en la validación de la herramienta con casos de estudio relevantes al área. Los casos de estudio utilizan PQR reales de Buenos Aires. Se ilustró el uso de la herramienta con dos conjuntos de datos diferentes, analizando los servicios y zonas más afectadas en la ciudad.

La segunda herramienta desarrollada se presentó en el Capítulo 5, donde se definieron diferentes técnicas que permiten la obtención y análisis de información para el análisis de interacción gobierno-ciudadano, en particular en Twitter. Se estableció una forma de recolectar tweets de cuentas de gobierno y las correspondientes respuestas de los ciudadanos, y se definieron las técnicas de minería de datos y de visualización que se pueden utilizar para su análisis. Se utilizó una técnica de análisis de sentimiento y diferentes visualizaciones acordes al procesamiento que pueden realizarse con la información recolectada. Las visualizaciones fueron gráficos de serie de tiempo para el análisis de actividad de la cuenta, además de las diferentes interacciones recibidas y emociones identificadas;

diagramas de Euler para estudiar el uso de recursos multimedia; diagramas de Hasse extendido para visualizar las interacciones ciudadanas; diagramas de Sankey para observar la relación entre el uso de recursos multimedia y las respuestas obtenidas por los ciudadanos; gráficos de radar y diagramas de burbujas para estudiar las emociones identificadas en los comentarios. También se lograron cumplir los objetivos planteados de implementar un prototipo del sistema propuesto y validarlo con casos de estudio relevantes. El diseño e implementación de dicha plataforma se presentaron en las secciones 5.4 y 5.5. Los casos de estudio se centraron en analizar las interacciones gobierno-ciudadano en diferentes países de Latinoamérica durante un mes (Sección 5.6) y de cinco secretarías de México durante nueve meses (Sección 5.7), utilizando diferentes iteraciones de la herramienta.

Finalmente, la tercera se explicó en el Capítulo 6, que busca facilitar la minería de datos heterogéneos, y permitir identificar y recuperar un conjunto reducido de datos relevantes. La solución propuesta es un sistema que permite recuperar datos de diferentes fuentes, actualmente de Google Noticias, Reddit y Twitter, y utiliza modelos de clasificación y ciclos de aprendizaje activo para definir los ítems de información relevantes para el usuario. Estas herramientas pueden formar parte de una infraestructura de software de utilidad para funcionarios públicos y usuarios expertos de cualquier tipo. Los objetivos planteados en la Sección 1.2 también se cumplieron: el desarrollo del prototipo, cuyo diseño e implementación se indican en las secciones 6.4 y 6.5, y la validación mediante casos de estudio. Los usuarios en ambos casos fueron estudiantes de magister en Desarrollo Internacional en Saint Mary's University (Halifax, Canadá), y el detalle de ambos pueden verse en las secciones 6.6 y 6.7.

7.2. Discusión

A lo largo de esta tesis se presentaron tres herramientas de utilidad para funcionarios públicos y usuarios expertos en general. Cada herramienta contiene un conjunto de técnicas que pueden ser parte de una infraestructura de software. A su vez, cada plataforma provee una serie de facilidades para el análisis de diferente tipo de datos. A continuación, se describirán las herramientas desarrolladas indicando brevemente sus diferencias con otros trabajos ya analizados en el Capítulo 3.

La herramienta descrita en el Capítulo 4 ofrece funcionalidades para el análisis de PQR ciudadanos. De las herramientas identificadas en los trabajos relacionados que utilizan da-

tos de este estilo (BJB⁺18; FPB⁺17), las más parecidas a la plataforma desarrollada hacen uso de visualizaciones como mapas de calor (KH14a) y temáticos, y aplican algoritmos de agrupamiento sobre los temas de los reportes (DHS⁺17). Sin embargo, la plataforma desarrollada por la tesista utiliza algoritmos de agrupación sobre la ubicación geográfica de las PQR, que enriquecen las visualizaciones utilizadas, que son mapas de calor y de agrupaciones. La herramienta facilita distinguir distintas concentraciones de reportes a lo largo de diferentes zonas geográficas, haciendo más simple el trabajo de reconocer las áreas más afectadas. Esto se traduce a una ayuda en la planificación logística a fin de atender las demandas más urgentes de manera eficiente. Sus limitaciones están asociadas a la aplicación de algoritmos que deben ejecutarse en *background* y periódicamente. Es importante el análisis del tiempo adecuado para su ejecución, considerando la fluctuación entre la generación y resolución de los reportes. También debe tenerse en cuenta la posibilidad de cambiar estos algoritmos por otros más eficientes de la misma índole, más recientes en la literatura (MHA17; JJ19).

En los Capítulos 5 y 6 se presentan herramientas que permiten obtener información de diferentes fuentes. La primera plataforma recupera información de Twitter o de periódicos digitales, y ofrece una variedad de visualizaciones a fin de estudiar el comportamiento en cuentas de gobierno y la interacción con los ciudadanos, incluyendo las emociones detectadas en los comentarios. Otros trabajos parecidos identificados en el Capítulo 3 realizan un análisis estadístico del archivo histórico (AA15) o usan técnicas de minería de datos y modelado de simulación, y producen una serie de métricas sobre interés y aceptación, entre otros (CGFL10; CL12; CL15). Por otro lado, la herramienta desarrollada por la tesista permite obtener dinámicamente los últimos tweets de una cuenta oficial de Twitter, y generar diversas visualizaciones que posibilitan hacer foco en los niveles de actividad del gobierno, examinar las palabras usadas y los temas discutidos más destacados, visualizar la estrategia de comunicación adoptada, y explorar las interacciones y emociones de los ciudadanos detectadas en los comentarios. Esto ayuda a identificar cómo los ciudadanos perciben las decisiones gubernamentales o las políticas públicas, y permite reconocer la relación entre estrategias de comunicación y respuesta ciudadana. Sus limitaciones están relacionadas a procesamientos adicionales que podrían agregarse a la herramienta para perfeccionar los datos analizados, como es la adopción de un algoritmo de análisis de sentimiento más sofisticado, funcionalidades para detectar errores ortográficos, ambigüedades, *trolling* y *spamming*.

La plataforma descrita en el Capítulo 6 presenta una metodología para la recuperación de información relevante de diversas fuentes sobre temáticas específicas, que no se ha encontrado en el trabajo relacionado aplicado al área de participación ciudadana. La herramienta permite la creación de modelos de clasificación, la recuperación de ítems de información de Google Noticias, Reddit y Twitter, y el entrenamiento de los modelos mediante la clasificación manual de un número reducido de ítems. Así mismo, ofrece una visualización del comportamiento del modelo, y dos opciones de exportación de documentos relevantes. La plataforma resulta de ayuda para reducir los tiempos de identificación de información relevante dentro del gran volumen de información que se puede recuperar de la web. Las limitaciones están relacionadas a las APIs utilizadas, que resulta en largos tiempos de espera cuando se realizan búsquedas por palabras clave. También se requieren nuevas funcionalidades, como la opción de eliminar las noticias relacionadas a determinadas búsquedas, permitir la importación de documentos, e incluir nuevas fuentes de información para realizar nuevas búsquedas.

7.3. Trabajo futuro

Durante el desarrollo de las diferentes herramientas que surgieron como parte de esta tesis se identificaron diferentes limitaciones, mejoras y agregados que quedaron a realizar como trabajo a futuro. Entre ellas, y en particular para la herramienta del Capítulo 4, la inclusión de algoritmos de agrupamiento que tomen en cuenta la dimensión temporal de los reportes de los ciudadanos, y de fuentes adicionales de datos como lo son las opiniones ciudadanas en redes sociales o periódicos digitales, aquellos específicamente asociados a los reclamos dentro de un municipio. Esto podría derivar en la aplicación de búsqueda basada en temas, reconocimiento de entidades nombradas y análisis de sentimiento, entre otras. Para la plataforma del Capítulo 5, se dejó pendiente la aplicación de algoritmos de análisis de sentimiento más sofisticados, y las mejoras en el procesamiento de lenguaje natural, incluyendo funcionalidades para identificar errores ortográficos, ironía, ambigüedades y regionalismos, entre otros. La posibilidad de detectar *trolling* y *spamming* serían funcionalidades extras que ayudarían a limitar la influencia del ruido en el análisis. Así mismo, se consideraron tareas tales como permitir la recuperación de tweets utilizando hashtags o palabras clave, y utilizar la herramienta en otras redes sociales como Facebook e Instagram, lo que implica la adaptación e implementación teniendo en

cuenta los mecanismos propios de interacción entre usuarios y las conexiones a las APIs correspondientes. Por su parte, para la plataforma del Capítulo 6 quedaron pendientes las tareas de eliminar las noticias asociadas a determinadas búsquedas (es decir, aquellas obtenidas utilizando ciertas palabras clave), agregar la opción de importar documentos, y la posibilidad de obtener ítems de información de otras fuentes, como pueden ser las redes sociales anteriormente mencionadas.

Otra línea de trabajo futuro es integrar y empaquetar las herramientas definidas, tanto para la minería, análisis, agrupamiento y visualización de datos, en un modelo de infraestructura de software. Esta tarea se puede realizar en base a la definición y diagramas de casos de uso y de componentes (definidos en los Capítulos 4, 5 y 6), identificando claramente los usuarios, los casos de uso y los componentes en común.

Adicionalmente, se espera poder abordar en el futuro el desarrollo de una herramienta que permita identificar las diferentes temáticas que pueden llegar a discutirse en hilos de comentarios presentes en determinadas redes sociales (como Twitter y Reddit) y, en particular, en periódicos digitales. Teniendo en cuenta que generalmente los usuarios suelen comentar e interactuar entre sí exponiendo su punto de vista, y que es natural la ramificación de las temáticas abordadas, el propósito de la herramienta es reconocer estas temáticas y las emociones que pueden identificarse en los comentarios.

7.4. Publicaciones

Parte de los resultados obtenidos durante el desarrollo de esta tesis fueron publicados en conferencias nacionales, internacionales y revistas de la especialidad. A continuación, se detallarán dichos trabajos.

En el trabajo «*CitymisVis: a tool for the visual analysis and exploration of citizen requests and complaints*» (HMCM17), publicado en la *10th International Conference on Theory and Practice of Electronic Governance* (Conferencia Internacional sobre Teoría y Práctica de la Gobernanza Electrónica) se presentan los resultados de utilizar la herramienta de visualización de PQR ciudadanos sobre los datos provistos por Mismática. La información de este trabajo se corresponde con la plataforma descrita en el Capítulo 4, y en particular con el caso de estudio de la Sección 4.6.

El trabajo «*Examining government-citizen interactions on Twitter using visual and sentiment analysis*» (HEMJ18), publicado en la *19th Annual International Conference*

on *Digital Government Research: Governance in the Data Age* (Conferencia Internacional Anual sobre Investigación del Gobierno Digital: Gobernanza en la era de los datos) presenta una metodología que comprende un conjunto de diferentes técnicas de visualización para analizar las interacciones gobierno-ciudadano. El análisis se enfocó en las interacciones que tuvieron lugar en Twitter, principalmente a través de las cuentas oficiales de los ministerios de cinco países de Latinoamérica. Este trabajo se corresponde con la herramienta que se presenta en el Capítulo 5, y con el caso de estudio de la Sección 5.6.

La publicación «*Modelo y plataforma de participación ciudadana para gobiernos municipales*» (HEM19), siendo un capítulo del libro «*Soluciones de gobernanza electrónica para la participación ciudadana*», realiza un análisis de los temas de interés de participación ciudadana a nivel municipal basado en una revisión de literatura. Se propone un modelo conceptual de participación ciudadana para gobiernos municipales, que luego es ilustrado con un caso de estudio utilizando los datos de la iniciativa «Gestión Colaborativa» de la Ciudad de Buenos Aires, visualizados a través de la herramienta detallada en el Capítulo 4. Los resultados del caso de estudio se presentan en esta tesis en la Sección 4.7.

El trabajo «*Analyzing and visualizing government-citizen interactions on Twitter to support public policy-making*» (HEMJ20), publicado en la revista «*Digital Government: Research and Practice*» (Gobierno Digital: Investigación y Práctica) presenta los resultados del análisis de interacciones gobierno-ciudadano en Twitter de cinco secretarías de México, aplicando diferentes técnicas de visualización y minería de opinión. Este trabajo utiliza la última versión de la plataforma del Capítulo 5. En este caso, los resultados del trabajo son aquellos expuestos en el caso de estudio de la Sección 5.7.

El trabajo «*Exploring the state of practice in e-Participation - technologies, trends and policy recommendations*» (actualmente en revisión) resume el estado del arte en el área de e-Participación mediante una revisión sistemática de artículos que presentan iniciativas y plataformas desarrolladas para la participación ciudadana, poniendo atención en las tecnologías utilizadas. Además de centrarse en el uso de técnicas y cómo se relacionan con los niveles y áreas de e-Participación, se ofrecen un conjunto de recomendaciones generales para el uso eficaz de las técnicas identificadas en los trabajos a fin de guiar el desarrollo de una nueva iniciativa de e-Participación. Este trabajo se corresponde con el contenido del Capítulo 3 y del Apéndice A.

Apéndice A

Clasificación de iniciativas de participación ciudadana

En este apéndice se presenta la clasificación realizada de los trabajos identificados en la literatura (Sección 3.2), que presentaran alguna iniciativa o plataforma de e-Participación. Estas plataformas fueron clasificadas teniendo en cuenta su nivel y área de participación y quién dio origen a la iniciativa (Sección 2.1), qué técnicas utiliza (Sección 2.3.2) y qué tipo de validación se realizó (Sección 3.4).

La Tabla A.1 presenta la clasificación de todas las iniciativas. En las columnas de «Áreas», las celdas marcadas con × indican el área principal y las marcadas con + el área secundaria. Algunas iniciativas no tienen un área secundaria. Cada técnica utilizada se marca con ×, sin indicar orden o preferencia. Los orígenes se identifican como:

- IN: liderados por la investigación.
- IG: liderados por la investigación con ayuda gubernamental.
- GO: dirigidos por el gobierno.
- GI: dirigidos por el gobierno con participación de investigadores.
- SC: liderados por una sociedad civil.

Finalmente, las validaciones se identifican como:

- CL: validación cualitativa.
- CT: validación cuantitativa.
- CC: validación cuantitativa y cualitativa.

| Nivel | Referencia | Origen | Áreas | | | | | | | | | | Técnicas | | | | | | | | | | | |
|----------------|------------|--------|--------------------------|--------------------------|----------|---------|-------------------|--------------|----------|-----------|---------------------------|----------|----------|--------------------|---------------|--------------------|------------------|--------------------|--------------|-----|------------------------|------------------------|-----------------|-----------------------------|
| | | | Provisión de información | Construcción comunitaria | Consulta | Campaña | Campaña electoral | Deliberación | Discurso | Mediación | Planificación territorial | Encuesta | Votación | Realidad aumentada | Crowdsourcing | Analítica de datos | Minería de datos | Minería de opinión | Gamificación | GIS | Tecnologías semánticas | Modelado de simulación | Visualizaciones | Visualización de argumentos |
| e-Consulta | (SPSB17) | IN | | | × | | | | | | | | | | | | | × | | | | | | - |
| | (PP17) | IG | | | × | | | | | | | | × | × | × | | | × | | | | × | | - |
| | (SFL+17) | IN | | | + | | | × | | | | | | × | | × | | | × | | | × | | - |
| | (SFB+17) | IN | | | + | | | × | | | | | | × | | × | | | × | | | × | | - |
| | (PL18) | IG | | | + | | | | | | × | | | × | | | | × | | | | × | | CT |
| | (QG18) | SC | | | × | | | | | | | | | × | | | | | | | | | × | CT |
| | (BGK+18) | IG | | | × | | | | | | | | × | × | | | | × | | | | × | | CC |
| | (DHS+17) | GI | | | × | | | | | | | | | | × | | | × | | | | × | | - |
| | (HDS17) | GI | | | × | | | | | | | | | | × | | | × | | | | × | | CT |
| | (SAAG18) | IN | | | × | | | | | | | | | × | | | | | | | | | | - |
| (FJ19) | GO | | | | | | | | | × | | | | | | | | × | | | | | - | |
| e-Colaboración | (EA) | IN | | | × | | | | | | | | × | | | | | × | | | | | | - |
| | (TDPT11) | IN | | | × | | | | | | | | | | | | | | | | | | × | CC |
| | (Ber12) | SC | | | | | | | | × | | | × | | | | × | | | | | × | | CL |
| | (AM12) | IN | | | × | | | | | | | | × | × | | × | | × | × | | | × | | CL |
| | (SWA+12) | IG | | | × | | | | | | | | × | × | | | | | | | | | | - |
| | (CFE13) | IN | | × | | | | | | | | | | | | | | | | | | | | - |
| | (CME+13) | IN | | | × | | | | | | | | | | | × | | × | | | | × | | - |
| | (POB18) | IN | | | | | | × | | | | | | × | | × | | × | | | | × | | CL |
| | (ENV14) | IN | | | | | | × | | | | | | × | | × | | | | | | | | - |
| | (KH14a) | IN | | | | × | | | | | | | × | | | | | × | | | | × | | - |
| | (KLLS14) | IN | | | × | | | | | + | | | × | | | | | × | | × | | | | CC |
| | (CMG14) | IN | | | × | | | | | | | | | | | × | | × | | | | × | | - |
| | (SG14) | IG | | | | × | | | | | | | | | | | | | | | | | | CT |
| | (LBE+14) | IN | | | × | | | | | + | | | | | | | × | × | | | | × | | CC |
| | (SCLD14) | IN | | | | | | | | | × | | | × | × | | × | | | | | × | | CC |
| | (SA15) | IN | | | | | | | | | | | | | | | | | | | | × | | CC |
| (TLSG15) | IN | | | × | | | | | | + | | | | | | | × | × | | | × | | CC | |
| (TL15) | IN | | | × | | | | | | + | | | | | | | × | × | | | × | | CC | |

Bibliografía

- [AA15] AL-AAMA, A. Y. The use of twitter to promote e-participation: connecting government and people. *International Journal of Web Based Communities* 11, 1 (2015), 73–96.
- [Age21] Agenda 21, Mar 2021. [Online; accessed 2. Mar. 2021].
- [AKLC18] ANDROUTSOPOULOU, A., KARACAPILIDIS, N., LOUKIS, E., AND CHARALABIDIS, Y. Combining technocrats’ expertise with public opinion through an innovative e-participation platform. *IEEE Transactions on Emerging Topics in Computing* (2018).
- [AM12] ANDREA, K., AND MEDENI, T. D. Ubipol approach to improve policy making-strenghts and weaknesses. In *Proceedings of the European, Mediterranean & Middle Eastern Conference on Information Systems* (2012).
- [ARA11] ALLEN, M., REGENBRECHT, H., AND ABBOTT, M. Smart-phone augmented reality for public participation in urban planning. In *Proceedings of the 23rd Australian computer-human interaction conference* (2011), pp. 11–20.
- [ARS⁺17] ALRYALAT, M. A. A., RANA, N. P., SAHU, G. P., DWIVEDI, Y. K., AND TAJVIDI, M. Use of social media in citizen-centric electronic government services: A literature analysis. *International Journal of Electronic Government Research (IJEGR)* 13, 3 (2017), 55–79.
- [BC12] BANNISTER, F., AND CONNOLLY, R. Forward to the past: Lessons for the future of e-government from the story so far. *Information Polity* 17, 3, 4 (2012), 211–226.
- [BCK03] BASS, L., CLEMENTS, P., AND KAZMAN, R. *Software architecture in practice*. Addison-Wesley Professional, 2003.

- [BEJ16] BERTOT, J., ESTEVEZ, E., AND JANOWSKI, T. Universal and contextualized public services: Digital public service innovation framework, 2016.
- [Ber06] BERKHIN, P. A survey of clustering data mining techniques. In *Grouping multidimensional data*. Springer, 2006, pp. 25–71.
- [Ber12] BERIX, M. Yutpa as a design tool for public participation. *AI & society* 27, 1 (2012), 165–172.
- [BFF15] BALAHADIA, F. F., FABITO, B. S., AND FERNANDO, M. C. G. E-participation: Incident mapping portal for local government units. In *2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNI-CEM)* (2015), IEEE, pp. 1–6.
- [BFR16] BIANCHINI, D., FOGLI, D., AND RAGAZZI, D. Tab sharing: A gamified tool for e-participation. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2016), pp. 294–295.
- [BGK⁺18] BOJOVIC, D., GIUPPONI, C., KLUG, H., MORPER-BUSCH, L., COJOCARU, G., AND SCHÖRGHOFER, R. An online platform supporting the analysis of water adaptation measures in the alps. *Journal of Environmental Planning and Management* 61, 2 (2018), 214–229.
- [BJB⁺18] BERNTZEN, L., JOHANNESSEN, M. R., BÖHM, S., WEBER, C., AND MORALES, R. Citizens as sensors: human sensors as a smart city data source. In *SMART 2018-The Seventh International Conference on Smart Systems, Devices and Technologies* (2018).
- [BM12] BENN, N., AND MACINTOSH, A. Policycommons—visualizing arguments in policy consultation. In *International Conference on Electronic Participation* (2012), Springer, pp. 61–72.
- [BML16] BOUDJELIDA, A., MELLOULI, S., AND LEE, J. Electronic citizens participation: Systematic review. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance* (2016), pp. 31–39.

- [BTRF12] BONSON, E., TORRES, L., ROYO, S., AND FLORES, F. Local e-government 2.0: Social media and corporate transparency in municipalities. *Government information quarterly* 29, 2 (2012), 123–132.
- [BYRN11] BAEZA-YATES, R. A., AND RIBEIRO-NETO, B. A. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., 2011.
- [CÁP13] CUCURULL, J., ÁLVARO, A., AND PUIGGALÍ, J. Myuniversity: E-participation and decision making for higher education. *Electronic Government and Electronic Participation-Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2036* (2013).
- [CFE13] CLEDOU, M. G., FERNANDES, S., AND ESTEVEZ, E. Welead: collaborative toolkit for learning, engaging and deciding. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance* (2013), pp. 378–379.
- [CGFL10] CHARALABIDIS, Y., GIONIS, G., FERRO, E., AND LOUKIS, E. Towards a systematic exploitation of web 2.0 and simulation modeling tools in public policy process. In *International Conference on Electronic Participation* (2010), Springer, pp. 1–12.
- [cit20] Citymis, Nov 2020. [Online; accessed 3. Nov. 2020].
- [CL12] CHARALABIDIS, Y., AND LOUKIS, E. Participative public policy making through multiple social media platforms utilization. *International Journal of Electronic Government Research (IJEGR)* 8, 3 (2012), 78–97.
- [CL15] CHARALABIDIS, Y., AND LOUKIS, E. Transforming government agencies' approach to eparticipation through multiple social media exploitation. In *From Information to Smart Society*. Springer, 2015, pp. 251–268.
- [CLA⁺14] CHARALABIDIS, Y., LOUKIS, E. N., ANDROUTSOPOULOU, A., KARKALETSIS, V., AND TRIANTAFILLOU, A. Passive crowdsourcing in government using social media. *Transforming Government: People, Process and Policy* 8, 2 (2014), 283–308.

- [CME⁺13] CHESÑEVAR, C., MAGUITMAN, A., ESTEVEZ, E., OSMAN, N., AND SIERRA, C. E2 participation: electronically empowering citizens for social innovation through agreement technologies. In *Proceedings of the 14th Annual International Conference on Digital Government Research* (2013), pp. 279–280.
- [CMG14] CHESÑEVAR, C. I., MAGUITMAN, A. G., AND GONZÁLEZ, M. P. Empowering citizens through opinion mining from twitter-based arguments. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance* (2014), pp. 275–278.
- [CVJ15] CARIC, A., VUKOVIC, M., AND JEVTIC, D. e-consultation: Automatic system for online consultations. In *2015 13th International Conference on Telecommunications (ConTEL)* (2015), IEEE, pp. 1–8.
- [dat19] The Data Spectrum: defining Shared & Closed, May 2019. [Online; accessed 3. Nov. 2020].
- [dat20] Datos Argentina, Oct 2020. [Online; accessed 1. Oct. 2020].
- [DB12] DESOUZA, K. C., AND BHAGWATWAR, A. Citizen apps to solve complex urban problems. *Journal of Urban Technology* 19, 3 (2012), 107–136.
- [dCPdSA19] DE CAMARGO PENTEADO, C. L., DOS SANTOS, M. B. P., AND ARAÚJO, R. D. P. A. Civil society and online citizen participation: A case study of the nossas cidades network. In *The Internet and Health in Brazil*. Springer, 2019, pp. 65–84.
- [DCPS08] DE CINDIO, F., PERABONI, C., AND SONNANTE, L. Improving citizens’ interactions in an e-deliberation environment. In *Proceedings of the working conference on Advanced visual interfaces* (2008), pp. 486–487.
- [DHS⁺17] DHINI, A., HARDAYA, I. S., SURJANDARI, I., ET AL. Clustering and visualization of community complaints and proposals using text mining and geographic information system. In *2017 3rd International Conference on Science in Information Technology (ICSITech)* (2017), IEEE, pp. 132–137.
- [DSN⁺11] DETERDING, S., SICART, M., NACKE, L., O’HARA, K., AND DIXON, D. Gamification. using game-design elements in non-gaming contexts. In

- CHI'11 extended abstracts on human factors in computing systems*. 2011, pp. 2425–2428.
- [EA] EKELIN, A., AND ANDERBERG, P. The augment project: Co-constructive accessibility mapping for supporting people in their realities. *Guest Editors*, 9.
- [EKS⁺96] ESTER, M., KRIEGEL, H.-P., SANDER, J., XU, X., ET AL. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (1996), vol. 96, pp. 226–231.
- [EMT11] ERGAZAKIS, K., METAXIOTIS, K., AND TSITSANIS, T. A state-of-the-art review of applied forms and areas, tools and technologies for e-participation. *International Journal of Electronic Government Research (IJEGR)* 7, 1 (2011), 1–19.
- [ENV14] EPSTEIN, D., NEWHART, M., AND VERNON, R. Not by technology alone: The “analog” aspects of online public engagement in policymaking. *Government Information Quarterly* 31, 2 (2014), 337–344.
- [FJ19] FATHEJALALI, A., AND JAIN, A. Mobile participation (mparticipation) in urban development: The experience of flashpoll app in berlin (germany). *Information Polity* 24, 2 (2019), 199–222.
- [FLL⁺06] FRASER, C., LIOTAS, N., LIPPA, B., MACH, M., MACINTOSH, F. M., MENTZAS, G., ROSENDAHL, A., SABOL, T., TAMBOURIS, E., TARABANIS, K., ET AL. net: Deliverable 5.1 report on current icts to enable participation.
- [FPB⁺17] FERREIRA, R. S., PRATA, R., BARBOSA, C. E., DE SOUZA, J. M., MORORO, V., AND COTTA, K. P. Transreport: collaborative supervision of the public transportation. In *Proceedings of the Symposium on Applied Computing* (2017), pp. 1808–1813.
- [FTA12] FEDOTOVA, O., TEIXEIRA, L., AND ALVELOS, H. E-participation in portugal: evaluation of government electronic platforms. *Procedia Technology* 5 (2012), 152–161.

- [GR13] GLIDDEN, J., AND RUSTON, S. Puzzled by policy: Providing new hope for e-participation at the eu level. *Electronic Government and Electronic Participation-Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2037* (2013).
- [Gra14] GRAY, J. Towards a genealogy of open data. In *The paper was given at the General Conference of the European Consortium for Political Research in Glasgow* (2014).
- [Gru93] GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge acquisition* 5, 2 (1993), 199–220.
- [HA14] HAND, D. J., AND ADAMS, N. M. Data mining. *Wiley StatsRef: Statistics Reference Online* (2014), 1–7.
- [Han07] HANZL, M. Information technology as a tool for public participation in urban planning: a review of experiments and potentials. *Design studies* 28, 3 (2007), 289–307.
- [HBML17] HE, G., BOAS, I., MOL, A. P., AND LU, Y. E-participation for environmental sustainability in transitional urban china. *Sustainability Science* 12, 2 (2017), 187–202.
- [HDS17] HARDAYA, I. S., DHINI, A., AND SURJANDARI, I. Application of text mining for classification of community complaints and proposals. In *2017 3rd International Conference on Science in Information Technology (ICSITech)* (2017), IEEE, pp. 144–149.
- [HEM19] HUBERT, R., ESTEVEZ, E., AND MAGUITMAN, A. Modelo y plataforma de participación ciudadana para gobiernos municipales. *Ana Maguitman Carlos Chesñevar* (2019).
- [HEMJ18] HUBERT, R. B., ESTEVEZ, E., MAGUITMAN, A., AND JANOWSKI, T. Examining government-citizen interactions on twitter using visual and sentiment analysis. In *Proceedings of the 19th annual international conference on digital government research: governance in the data age* (2018), pp. 1–10.

- [HEMJ20] HUBERT, R. B., ESTEVEZ, E., MAGUITMAN, A., AND JANOWSKI, T. Analyzing and visualizing government-citizen interactions on twitter to support public policy-making. *Digital Government: Research and Practice* 1, 2 (2020), 1–20.
- [HMCM17] HUBERT, R. B., MAGUITMAN, A. G., CHESÑEVAR, C. I., AND MALAMUD, M. A. Citymisvis: a tool for the visual analysis and exploration of citizen requests and complaints. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance* (2017), pp. 22–25.
- [Hon13] HONG, H. Government websites and social media’s influence on government-public relationships. *Public Relations Review* 39, 4 (2013), 346–356.
- [iap20] IAP2 IAP2 Spectrum | IAP2 Australasia, Nov 2020. [Online; accessed 6. Nov. 2020].
- [JH18] JANSSEN, M., AND HELBIG, N. Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly* 35, 4 (2018), S99–S105.
- [JJ19] JANG, J., AND JIANG, H. Dbscan: Towards fast and scalable density clustering. In *International Conference on Machine Learning* (2019), PMLR, pp. 3019–3029.
- [JMZ15] JANSSEN, M., MATHEUS, R., AND ZUIDERWIJK, A. Big and open linked data (bold) to create smart cities and citizens: Insights from smart energy and mobility cases. In *International Conference on Electronic Government* (2015), Springer, pp. 79–90.
- [JPWP15] JUNG, K., PARK, S. J., WU, W.-N., AND PARK, H. W. A webometric approach to policy analysis and management using exponential random graph models. *Quality & Quantity* 49, 2 (2015), 581–598.
- [JWT11] JOSHI, S., WANDHOEFER, T., AND THAMM, M. Rethinking governance via social networking: The case of direct vs. indirect stakeholder injection. In *Proceedings of the 5th International Conference on Theory and Practice of electronic Governance* (2011), pp. 277–280.

- [KH14a] KUBÁSEK, M., AND HŘEBÍČEK, J. Involving citizens into mapping of illegal landfills and other civic issues in the czech republic.
- [KH14b] KUKOVIC, S., AND HAČEK, M. The distribution of e-democracy and e-participation tools in slovenian municipalities. *World Political Science* 10, 1 (2014), 87–106.
- [Kin04] KING, S. *Dark Tower*. Scribner, 2004.
- [KLLS14] KHAN, Z., LUDLOW, D., LOIBL, W., AND SOOMRO, K. Ict enabled participatory urban planning and policy development. *Transforming Government: People, Process and Policy* (2014).
- [LBE⁺14] LEHNER, U., BALDAUF, M., ERANTI, V., REITBERGER, W., AND FRÖHLICH, P. Civic engagement meets pervasive gaming: towards long-term mobile participation. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 2014, pp. 1483–1488.
- [LDD19] LAFRANCE, F., DANIEL, S., AND DRAGIĆEVIĆ, S. Multidimensional web gis approach for citizen participation on urban evolution. *ISPRS International Journal of Geo-Information* 8, 6 (2019), 253.
- [Mac04] MACINTOSH, A. Characterizing e-participation in policy-making. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (2004), IEEE, pp. 10–pp.
- [Map18] MAPBOX. Clustering millions of points on a map with Supercluster. *Medium* (Jun 2018).
- [Med11] MEDAGLIA, R. eparticipation research: A longitudinal overview. In *International Conference on Electronic Participation* (2011), Springer, pp. 99–108.
- [MHA17] MCINNES, L., HEALY, J., AND ASTELS, S. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* 2, 11 (2017), 205.
- [mis20] Mismatica, Oct 2020. [Online; accessed 1. Oct. 2020].
- [MMD17] MARZOUKI, A., MELLOULI, S., AND DANIEL, S. Towards a context-based citizen participation approach: A literature review of citizen participation

- issues and a conceptual framework. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance* (2017), pp. 204–213.
- [MP14] MOISE, M., AND POPA, V. An integrated policy system (ips) for supporting policy making. *Studies in Informatics and Control* 23, 3 (2014), 278.
- [MT10] MOHAMMAD, S., AND TURNEY, P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (2010), pp. 26–34.
- [Mur10] MUREȘAN, M. E-governance-a challenge for the regional sustainable development in romania. In *International Conference on Electronic Government* (2010), Springer, pp. 238–250.
- [nbn20] NBN Gateway - National Biodiversity Network, Nov 2020. [Online; accessed 3. Nov. 2020].
- [nC⁺06] NET CONSORTIUM, D., ET AL. Introducing eparticipation. demo-net booklet series. vol. 1, 2006.
- [OAOM16] ONI, A. A., AYO, C., ONI, S., AND MBARIKA, V. Strategic framework for e-democracy development and sustainability. *Transforming Government: People, Process and Policy* (2016).
- [OG19] OLIVEIRA, C., AND GARCIA, A. C. Citizens' electronic participation: a systematic review of their challenges and how to overcome them. *International Journal of Web Based Communities* 15, 2 (2019), 123–150.
- [ope18] Open Government Data, Dec 2018. [Online; accessed 1. Oct. 2020].
- [Ope20] OPEN KNOWLEDGE OPEN DEFINITION GROUP. Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge, Sep 2020. [Online; accessed 1. Oct. 2020].
- [OW09] ONG, C.-S., AND WANG, S.-W. Managing citizen-initiated email contacts. *Government Information Quarterly* 26, 3 (2009), 498–504.

- [OW13] OLIVEIRA, G. H. M., AND WELCH, E. W. Social media use in local government: Linkage of technology, task, and organizational context. *Government Information Quarterly* 30, 4 (2013), 397–405.
- [PK08] PHANG, C. W., AND KANKANHALLI, A. A framework of ict exploitation for e-participation initiatives. *Communications of the ACM* 51, 12 (2008), 128–132.
- [PL⁺01] PEÑA-LÓPEZ, I., ET AL. Citizens as partners. oecd handbook on information, consultation and public participation in policy-making.
- [PL18] PEÑAFIEL, V., AND LAVÍN, J. M. Improving university decision making through e-participation. In *2018 International Conference on eDemocracy & eGovernment (ICEDEG)* (2018), IEEE, pp. 392–396.
- [Plu01] PLUTCHIK, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.
- [POB⁺12] PORWOL, L., O’DONOGHUE, P., BRESLIN, J., COUGHLAN, C., AND MULLIGAN, B. Social inclusion and digital divide: eparticipation dilemmas in municipalities. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (2012), pp. 389–392.
- [POB16] PORWOL, L., OJO, A., AND BRESLIN, J. G. An ontology for next generation e-participation initiatives. *Government Information Quarterly* 33, 3 (2016), 583–594.
- [POB18] PORWOL, L., OJO, A., AND BRESLIN, J. G. Social software infrastructure for e-participation. *Government Information Quarterly* 35, 4 (2018), S88–S98.
- [PP17] PÁNEK, J., AND PÁSZTO, V. Emotional mapping in local neighbourhood planning: case study of příbram, czech republic. *International Journal of E-Planning Research (IJEPR)* 6, 1 (2017), 1–22.
- [QG18] QUENTAL, C., AND GOUVEIA, L. B. E-consultation as a tool for participation in teachers’ unions. In *Developments and Advances in Intelligent Systems and Applications*. Springer, 2018, pp. 153–167.

- [RBM16] ROMERO, C. D. G., BARRIGA, J. K. D., AND MOLANO, J. I. R. Big data meaning in the architecture of iot for smart cities. In *International Conference on Data Mining and Big Data* (2016), Springer, pp. 457–465.
- [RBS⁺14] REINWALD, F., BERGER, M., STOIK, C., PLATZER, M., AND DAMYANOVIC, D. Augmented reality at the service of participatory urban planning and community informatics—a case study from vienna. *The Journal of Community Informatics* 10, 3 (2014).
- [red20] Homepage - Reddit, Nov 2020. [Online; accessed 4. Nov. 2020].
- [RFT18] REXHEPI, A., FILIPOSKA, S., AND TRAJKOVIK, V. Youth e-participation as a pillar of sustainable societies. *Journal of Cleaner Production* 174 (2018), 114–122.
- [RY15] ROYO, S., AND YETANO, A. “crowdsourcing” as a tool for e-participation: two experiences regarding co 2 emissions at municipal level. *Electronic Commerce Research* 15, 3 (2015), 323–348.
- [SA15] SAMEER, L., AND ABDELGHAFAR, H. The use of social network in enhancing e-rulemaking. In *European Conference on Digital Government* (2015), Academic Conferences International Limited, p. 254.
- [SAAG18] SOUSA, A. A., AGANTE, P., ABRANTES, S., AND GOUVEIA, L. B. Libe-ropinion: A web platform for public participation. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance* (2018), pp. 199–208.
- [SC19] SANTINI, R. M., AND CARVALHO, H. The rise of participatory despotism: a systematic review of online platforms for political engagement. *Journal of Information, Communication and Ethics in Society* (2019).
- [SCB⁺10] STYLIOU, G., CHRISTODOULAKIS, D., BESHARAT, J., VONITSANO, M.-A., KOTROTSOS, I., KOUMPOURI, A., AND STAMO, S. Public opinion mining for governmental decisions. *Electronic Journal of e-Government* 8, 2 (2010), 202.
- [SCLD14] SPILIOPOULOU, L., CHARALABIDIS, Y., LOUKIS, E. N., AND DIAMANTOPOULOU, V. A framework for advanced social media exploitation in

government for crowdsourcing. *Transforming Government: People, Process and Policy* (2014).

- [SE15] STEIBEL, F., AND ESTEVEZ, E. Designing web 2.0 tools for online public consultation. In *Impact of information society research in the global south*. Springer, Singapore, 2015, pp. 243–263.
- [Set09] SETTLES, B. Active learning literature survey. Tech. rep., University of Wisconsin-Madison Department of Computer Sciences, 2009.
- [SFB⁺17] SCHMITZ, P., FRANCESCONI, E., BATOUCHE, B., LANDERCY, S.-P., AND TOULY, V. Ontological models of legal contents and users’ activities for eu e-participation services. In *International Conference on Electronic Government and the Information Systems Perspective* (2017), Springer, pp. 99–114.
- [SFL⁺17] SCHMITZ, P., FRANCESCONI, E., LANDERCY, S.-P., BATOUCHE, B., AND TOULY, V. A knowledge organization system for e-participation in law-making. In *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law* (2017), pp. 245–248.
- [SFS11] SÆBØ, Ø., FLAK, L. S., AND SEIN, M. K. Understanding the dynamics in e-participation initiatives: Looking through the genre and stakeholder lenses. *Government Information Quarterly* 28, 3 (2011), 416–425.
- [SG14] SUSHA, I., AND GRÖNLUND, Å. Context clues for the stall of the citizens’ initiative: lessons for opening up e-participation development practice. *Government Information Quarterly* 31, 3 (2014), 454–465.
- [SKS15] SUN, P.-L., KU, C.-Y., AND SHIH, D.-H. An implementation framework for e-government 2.0. *Telematics and Informatics* 32, 3 (2015), 504–520.
- [sol20] SolicitudesBA, Jan 2020. [Online; accessed 1. Oct. 2020].
- [SPCG19] SANTAMARÍA-PHILCO, A., CERDÁ, J. H. C., AND GRAMAJE, M. C. P. Advances in e-participation: A perspective of last years. *IEEE Access* 7 (2019), 155894–155916.
- [SPSB17] SINCLAIR, A. J., PEIRSON-SMITH, T. J., AND BOERCHERS, M. Environmental assessments in the internet age: the role of e-governance and social

- media in creating platforms for meaningful participation. *Impact Assessment and Project Appraisal* 35, 2 (2017), 148–157.
- [SS08] SCHNEPF, D., AND STRASSER, S. Steps towards sustainable urban development in the city of narva, estonia? supporting e-governance processes through the e-community tool? a web-based communication and management platform. *International Journal of Environmental Technology and Management* 8, 1 (2008), 53–68.
- [SSS19] STEINBACH, M., SIEWEKE, J., AND SÜSS, S. The diffusion of e-participation in public administrations: A systematic literature review. *Journal of Organizational Computing and Electronic Commerce* 29, 2 (2019), 61–95.
- [SWA⁺12] SINGH, G., WALSH, C. S., ALATHUR, S., ILAVARASAN, P. V., AND GUPTA, M. Citizen participation and effectiveness of e-petition: Sutharyakeralam–india. *Transforming Government: People, Process and Policy* (2012).
- [TDMW09] TIROPANIS, T., DAVIS, H., MILLARD, D., AND WEAL, M. Semantic technologies for learning and teaching in the web 2.0 era. *IEEE Intelligent Systems* 24, 6 (2009), 49–53.
- [TDPT11] TAMBOURIS, E., DALAKIOURIDOU, E., PANOPOULOU, E., AND TARABANIS, K. Evaluation of an argument visualisation platform by experts and policy makers. In *International Conference on Electronic Participation* (2011), Springer, pp. 74–86.
- [TEB17] THIEL, S.-K., ERTIÖ, T. P., AND BALDAUF, M. Why so serious? the role of gamification on motivation and engagement in e-participation. *IXD&A* 35 (2017), 158–181.
- [the20] The ODI – Open Data Institute, Nov 2020. [Online; accessed 3. Nov. 2020].
- [Thi16a] THIEL, S.-K. Gamers in public participation: a boon or bane? influence of attitudes in gamified participation platforms. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (2016), pp. 229–240.

- [Thi16b] THIEL, S.-K. A review of introducing game elements to e-participation. In *2016 conference for E-democracy and open government (CeDEM)* (2016), IEEE, pp. 3–9.
- [TKT08] TAMBOURIS, E., KALAMPOKIS, E., AND TARABANIS, K. A survey of e-participation research projects in the european union. *International Journal of Electronic Business* 6, 6 (2008), 554–571.
- [TL15] THIEL, S.-K., AND LEHNER, U. Exploring the effects of game elements in m-participation. In *Proceedings of the 2015 British HCI Conference* (2015), pp. 65–73.
- [TLSG15] THIEL, S.-K., LEHNER, U., STÜRMER, T., AND GOSPODAREK, J. Insights from a m-participation prototype in the wild. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (2015), IEEE, pp. 166–171.
- [TMC⁺07] TAMBOURIS, E., MACINTOSH, A., COLEMAN, S., WIMMER, M., VEDEL, T., WESTHOLM, H., LIPPA, B., DALAKIOURIDOU, E., PARISOPOULOS, K., ROSE, J., ET AL. Introducing eparticipation. *DEMO-net The Democracy Network. DEMO-net booklet series 1* (2007).
- [Too19] TOOTS, M. Why e-participation systems fail: The case of estonia’s osale.ee. *Government Information Quarterly* 36, 3 (2019), 546–559.
- [TPP09] TEUFL, P., PAYER, U., AND PARYCEK, P. Automated analysis of e-participation data by utilizing associative networks, spreading activation and unsupervised learning. In *International Conference on Electronic Participation* (2009), Springer, pp. 139–150.
- [TRR16] THIEL, S.-K., REISINGER, M., AND RÖDERER, K. i’m too old for this!influence of age on perception of gamified public participation. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (2016), pp. 343–346.
- [TRRB19] THIEL, S.-K., REISINGER, M., RÖDERER, K., AND BALDAUF, M. Inclusive gamified participation: Who are we inviting and who becomes enga-

- ged? In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (2019).
- [twi20] Twitter - Sobre nosotros, Nov 2020. [Online; accessed 4. Nov. 2020].
- [UNE20] UN E-Government Survey 2016, Dec 2020. [Online; accessed 3. Dec. 2020].
- [VKIS16] VIDIASOVA, L., KACHURINA, P., IVANOV, S., AND SMITH, G. E-participation tools in science and business sphere implementation: The case of xpir-platform for participation in education policy. *Procedia Computer Science* 101 (2016), 398–406.
- [VKS⁺17] VOGIATZI, M., KERATIDIS, C., SCHINAS, M., DIPLARIS, S., YÜMLÜ, S., FORBES, P., PAPADOPOULOS, S., SYROPOULOU, P., APOSTOLIDIS, L., KOMPATSIARIS, I., ET AL. The step project: societal and political engagement of young people in environmental issues. In *International Conference on Internet Science* (2017), Springer, pp. 148–156.
- [VT18] VIDIASOVA, L., AND TENSINA, I. E-participation social effectiveness: case of “our petersburg” portal. In *International Conference on Electronic Governance and Open Society: Challenges in Eurasia* (2018), Springer, pp. 308–318.
- [WDB18] WIRTZ, B. W., DAISER, P., AND BINKOWSKA, B. E-participation: A strategic framework. *International Journal of Public Administration* 41, 1 (2018), 1–12.
- [WGJH13] WIMMER, M. A., GRIMM, R., JAHN, N., AND HAMPE, J. F. Mobile participation: exploring mobile tools in e-participation. In *International Conference on Electronic Participation* (2013), Springer, pp. 1–13.
- [Wim07] WIMMER, M. A. Ontology for an e-participation virtual resource centre. In *Proceedings of the 1st international conference on Theory and practice of electronic governance* (2007), pp. 89–98.
- [XL10] XENAKIS, A., AND LOUKIS, E. An investigation of the use of structured e-forum for enhancing e-participation in parliaments. *International Journal of Electronic Governance* 3, 2 (2010), 134–147.

- [ZLP09] ZISSIS, D., LEKKAS, D., AND PAPADOPOULOU, A.-E. Competent electronic participation channels in electronic democracy. *Electronic Journal of E-government* 7, 2 (2009).
- [ZOC18] ZOLOTOV, M. N., OLIVEIRA, T., AND CASTELEYN, S. Continued intention to use online participatory budgeting: The effect of empowerment and habit. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance* (2018), pp. 209–216.