

**Análisis de la Calidad de Microdatos Extraídos de Twitter:  
Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y  
Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá  
2023**

**Arley Felipe Vera Yate**



**Pontificia Universidad Javeriana  
Facultad de Comunicación y Lenguaje  
Departamento de Ciencia información  
Carrera de Ciencia informacion Bibliotecologia y Archivistica  
Bogotá D.C  
Junio de 2024**

**Análisis de la Calidad de Microdatos Extraídos de Twitter:  
Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y  
Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá  
2023**

**Arley Felipe Vera Yate**

**Offray Vladimir Luna Cárdenas**  
Director de trabajo de grado



**Pontificia Universidad Javeriana  
Facultad de Comunicación y Lenguaje  
Departamento de Ciencia información  
Carrera de Ciencia informacion Bibliotecologia y Archivistica  
Bogotá D.C  
Junio de 2024**



# Contents

Problemática . . . . .	4
Justificación . . . . .	6
Estado del arte . . . . .	7
Objetivos . . . . .	11
Objetivo general . . . . .	11
Objetivos específicos . . . . .	11
Marco teórico . . . . .	12
Enfoque metodológico . . . . .	13
Desplegando un entorno de investigación reproducible . . . . .	15
Scoop . . . . .	15
Fossil y ChiselApp . . . . .	16
Glamorous Toolkit y Pharo . . . . .	17
Grafoscopio . . . . .	18
WindTerm . . . . .	20
Zettlr . . . . .	20
Pandoc . . . . .	21
Tectonic . . . . .	22
Eisvogel Template . . . . .	22
LiteXL . . . . .	23
Hypothesis . . . . .	24
Flameshot . . . . .	25
Publicación reproducible. . . . .	26
Adquisición de datos: Herramientas de Scraping . . . . .	35
Apify . . . . .	35
Twitter Scraper . . . . .	41
Squawker . . . . .	45
Análisis de la calidad de los microdatos extraídos . . . . .	59
Prerrequisitos . . . . .	59
Conclusiones y recomendaciones . . . . .	87
Bibliografía . . . . .	90

Anexos . . . . .	93
Tesis pregrado Felipe Vera: FileLocator alias tutor . . . . .	95
Exploración recursiva de estructuras de datos . . . . .	97
Uso de chatGPT . . . . .	103
Diccionarios Pharo Smalltalk . . . . .	104
Merge conflicts . . . . .	105

Este estudio se centra en examinar la calidad de los datos extraídos de las redes sociales, particularmente de Twitter (ahora denominada “X” y que a lo largo de esta tesis se denominará Twitter/X), en el contexto de las campañas políticas de las elecciones regionales del año 2023 en Bogotá. La información política que Gustavo Bolívar y Juan Daniel Oviedo publicaron en Twitter/X a lo largo de su campaña será analizada mediante minería de textos. El objetivo de este estudio es investigar herramientas o métodos de scraping para recopilar datos de la red social Twitter/X, por lo que este proyecto utiliza herramientas de análisis de datos para comprender cuánta información podemos extraer de los microdatos del discurso político en contextos digitales, particularmente en Twitter/X. Esto lo hace novedoso y pertinente para los campos de las ciencias de la información, las bibliotecas y los archivos. Los hallazgos del proyecto pueden ayudar a comprender los datos del discurso emitido por candidatos como manera de formentar maneras de participación política novedosas en contextos discursivos mediados digitalmente, que constituyen buena parte de la manera en que ciudadanos y candidatos se comunican contemporáneamente.

## Problemática

El crecimiento de la disponibilidad y accesibilidad de la información se denomina proliferación de la información (Méndez, 2019). Gracias a la tecnología y a la facilidad de compartir contenidos online, la información se produce y difunde a un ritmo cada vez mayor en la era digital. Además, la facilidad con la que se difunde la información en línea puede ser perjudicial para la seguridad y privacidad de las personas.

Para comunicar a las personas y permitirles compartir conocimientos y contenidos digitales, se desarrollaron plataformas en línea en la década de 1990, así surgieron las redes sociales (Grapsas, 2017) y entre ellas una de la más conocidas y reconocida en el mundo se llama Twitter/X. Esta fue desarrollada por Jack Dorsey, Biz Stone y Evan Williams en 2006 y puesto a disposición del público en general en 2007 (Marketing Zone Icesi, 2023).

Twitter, ahora conocido como X, fue desarrollado como una plataforma de microblogging que permitía a los usuarios publicar “tweets” de hasta 140 caracteres, que son mensajes de texto breves. La capacidad de intercambiar mensajes directos, transmitir videos en vivo, realizar encuestas e incluir imágenes y videos en tweets son solo algunas de las nuevas capacidades que Twitter/X ha desarrollado y agregado a través del tiempo (Marketing Zone Icesi, 2023). Desde que Elon Musk compró Twitter en octubre de 2022, se han presentado varios ajustes polémicos, entre ellos se destacan que a partir del 9 de febrero del 2023 Twitter realiza un cobro por el acceso a su API, que anteriormente estaba disponible para los desarrolladores e investigadores de forma gratuita al menos para bajos volúmenes de datos. (Efe et al., 2023). Aunque el precio aún no se ha revelado, esta medida tuvo un impacto en los investigadores y desarrolladores. Estas modificaciones representan el plan de Musk para gestionar y comercializar la plataforma e indican un nuevo rumbo para Twitter/X (Efe et al., 2023).

Por otro lado, estas plataformas brindan a los políticos y a los partidos que representan una forma rápida y asequible de contactar a una audiencia considerable en Internet y llegar a los votantes (Vlaicu, 2022), por lo que, en los últimos años la publicidad política en las redes sociales ha crecido, presentándose diversos tipos de contenido como la propaganda política en las redes sociales, incluidos memes, vídeos, anuncios pagados, publicaciones y noticias falsas, información que podrían influir en la opinión pública. (Vlaicu, 2022)

El caso de la empresa Cambridge Analytica se demostró el riesgo de utilizar las redes sociales en el ámbito político. Para dar contexto, esta empresa, sin autorización de millones de usuarios de la red social Facebook, recopiló y utilizó sus datos personales durante las elecciones presidenciales de EE. UU., con el fin de crear perfiles psicológicos detallados y personalizados de los votantes, para luego, utilizarlos para influir en su comportamiento electoral a través de publicidad altamente dirigida (Fernández, 2018). Este caso llevó a un mayor escrutinio de la privacidad de los datos de los usuarios de Facebook y a cambios significativos en la forma en que las empresas de tecnología manejan los datos

personales de los usuarios (Fernández, 2018).

En Colombia, diferentes políticos pueden utilizar las redes sociales en su beneficio, desde partidos y/o candidatos con el fin de buscar una ventaja en las elecciones con el interés de conseguir un impacto en el mundo de la política y las decisiones tomadas por el gobierno. Para las elecciones regionales de la alcaldía de Bogotá, se proporcionó la lista de potenciales candidatos junto con los enfoques políticos de los distintos partidos y movimientos. Cada candidato, según un comentario de Rcn (2023), se presenta como poseedor de los conocimientos, talentos y experiencia necesarios para mejorar la ciudad capitalina. Sin embargo, es importante destacar que esta afirmación refleja una opinión expresada por un medio periodístico y no necesariamente constituye un hecho verificable. En el año 2023 para la alcaldía de Bogotá se postularon como candidatos Juan Daniel Oviedo y Gustavo Bolívar, los cuales han sido elegidos como sujetos de esta investigación debido a su gran impacto en las redes sociales. Luego de ocupar el cargo de Director General del Departamento Administrativo Nacional de Estadística entre 2018 y 2022, Juan Daniel Oviedo se dio a conocer en la política nacional de Colombia. Es economista de la Universidad del Rosario y tiene un doctorado en economía en la universidad francesa de Toulouse (Doria et al., 2023). Oviedo se postula para la alcaldía de Bogotá como independiente y sin ningún tipo de alianza. Su rival, político y escritor de narconovelas de Colombia Gustavo Bolívar. Se dio a conocer como senador oponiéndose al uribismo y apoyando la campaña de Gustavo Petro desde 2018. Tiene un capital electoral entre los jóvenes que salieron a protestar en el 2021. Actualmente, le juega en contra el hecho de estar señalado de promover los desmanes de las últimas protestas, además de no contar con el conocimiento sobre la ciudad (Doria et al., 2023).

En este estudio, se destaca la importancia de la investigación reproducible y la narrativa de datos para garantizar la transparencia, verificabilidad y validez de los hallazgos. La reproducibilidad permite a otros investigadores verificar y validar los resultados, fortaleciendo la credibilidad de los estudios. La narrativa de datos proporciona un marco coherente para interpretar los resultados y comunicarlos efectivamente. Un recurso valioso que ilustra estos principios es “The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences” (Kitzes, Turek, & Deniz, 2018), que presenta 31 estudios de caso sobre flujos de trabajo de investigación reproducible, subrayando la importancia de herramientas y prácticas específicas para la investigación científica reproducible.

Al integrar estos elementos en la formulación del problema, se busca establecer un enfoque metodológico sólido para recopilar, analizar y presentar los datos de manera coherente y transparente. Esto es crucial para abordar la complejidad de los datos extraídos de Twitter/X sobre los perfiles de los candidatos a la alcaldía de Bogotá, asegurando un proceso de investigación riguroso y confiable para comprender la dinámica política en este contexto específico.

¿Cuál es la calidad de los microdatos que se puede extraer de esa red social mediante distintas fuentes?



## Justificación

El proyecto de análisis de microdatos de Gustavo Bolívar y Juan Daniel Oviedo sobre el impacto político en Twitter/X por medio de la extracción de datos durante su campaña a la alcaldía de Bogotá en el año 2023 demuestra características innovadoras de suma significación en términos de procesos y resultados esperados, entendiendo estos últimos como el impacto de la investigación. Además, hace uso de dos dominios de investigación que son pertinentes al campo de las ciencias de la información: biblioteca y archivística.

Para empezar, el proyecto está relacionado con el eje de estudio “Organización y gestión de la información y el conocimiento”. Nuestro objetivo es analizar la calidad de los microdatos extraídos de Twitter/X durante la candidatura a la alcaldía de Bogotá, utilizando técnicas de minería de datos. Con el uso de esta metodología, podemos evaluar la precisión y la fiabilidad de la información difundida en la plataforma y producir modelos útiles que pueden aplicarse para mejorar la toma de decisiones políticas y estratégicas. También promueve las Ciencias de la Información mediante el uso y la mejora del análisis de datos y en un entorno político y social particular.

En segundo lugar, el proyecto se encuadra en el eje de investigación “Información y conocimiento: usos, consumo y apropiación social”. Para comprender, analizar y explicar los fenómenos que impactan a la sociedad en términos de información y conocimiento, el análisis del impacto político Twitter/X durante su candidatura a la alcaldía de Bogotá pretende precisamente hacerlo. Para sugerir proyectos, programas y políticas que fomenten el desarrollo constructivo de la sociedad, es importante examinar cómo los usuarios de Twitter/X utilizan, consumen la información política socialmente apropiada. Este método nos ayuda a comprender cómo las redes sociales afectan los procesos políticos y electorales, la participación ciudadana. El proyecto también produce conocimientos y experiencia que avanzan en el estado del arte en las Ciencias de la Información y campos asociados. La comprensión de la calidad de los datos podemos extraer de los microdatos del discurso político en contextos digitales, particularmente en Twitter/X, se amplía mediante la aplicación de herramientas de análisis de datos de vanguardia, lo cual es crucial para la era de la información y la comunicación.

## Estado del arte

En el artículo “Uso de las redes sociales en diplomacia, política y relaciones internacionales. Análisis de la información publicada en las versiones online de dos periódicos españoles: “el país” y “la vanguardia”“. Indica un creciente interés en el uso de las redes sociales con fines políticos y diplomáticos. En el artículo se destaca el enfoque de las dos revistas adoptadas sobre este tema, destacando su importancia para la democracia y la transparencia (R. Cela et al., 2019). El periódico español “El País” explica este tema desde 2010 y adopta un enfoque más cuantitativo y multimedia, mientras que “La Vanguardia” ha mostrado interés más recientemente y se centra en cuestiones diplomáticas y relaciones entre líderes mundiales (R. Cela et al., 2019). El artículo destaca los efectos beneficiosos de las redes sociales en la política y al mismo tiempo crean conciencia sobre los peligros potenciales y la necesidad de un uso seguro. Con su uso generalizado, Twitter se ha convertido en la red política más popular en todo el mundo, apoyando el impacto en la formación de la opinión pública y la presencia de los líderes mundiales (Abdullah et al., 2022). El artículo “Examen de la relación entre los factores que influyen en la búsqueda de información política a través de los medios sociales entre los jóvenes de Malasia” analizó los factores que influyen en el comportamiento de los jóvenes de Malasia cuando buscan información política, enfatizando el creciente impacto de las redes sociales. Los hallazgos mostraron que una serie de características, incluida la utilidad, eficacia y experiencia de las redes sociales, tenían un impacto significativo en la forma en que los jóvenes encuentran información política (Abdullah et al., 2022). Específicamente, se descubrió que una mayor participación en las redes sociales estaba relacionada con un mayor comportamiento de búsqueda de información política. La política debería beneficiarse enormemente de estos hallazgos, que indican que las redes sociales pueden desempeñar un papel importante a la hora de involucrar a los jóvenes en el proceso político. Como resultado, el estudio enfatiza lo importante que es tener en cuenta las redes sociales al desarrollar estrategias para la participación política de los jóvenes en Malasia (Abdullah et al., 2022).

En el artículo “Fake news en tiempos de posverdad. Análisis de informaciones falsas publicadas en Facebook durante procesos políticos en Brasil y México 2018. Fake news in post-truth times” tienen como objetivo caracterizar los patrones discursivos de las noticias falsas publicadas en Facebook durante los procesos políticos en México en 2018, y también identificar dichos patrones discursivos. Se ha empleado una muestra de noticias falsas difundidas en Facebook durante los periodos de campaña presidencial en México, que previamente fueron identificadas por sitios de factchecking seleccionados (García y Gómez, 2022). El estudio se centra en el uso de la red social Facebook de los cuatro candidatos a la presidencia de México durante el periodo electoral de 2018 (García y Gómez, 2022). Las redes sociales, especialmente Facebook, fueron cruciales para la difusión de noticias políticas en Brasil a lo largo de los procesos políticos de 2018. La difusión de noticias falsas afectó los resultados electorales y la opinión pública (García y Gómez, 2022). La falta de conexiones y el uso de formatos similares de las publicaciones habituales en las redes sociales, como gráficos cautivadores y

textos atractivos, son características de estas noticias falsas (García y Gómez, 2022). Se utilizaron imágenes cargadas de emociones para provocar respuestas contundentes de las personas y fomentar la difusión del contenido. Se han descubierto una serie de tácticas que se emplean para aumentar la visibilidad de las noticias falsas en Facebook. Este tipo de propaganda, por ejemplo, se ha promovido a través de grupos y sitios especializados, lo que amplifica su efecto, para mejorar la exposición de las publicaciones y la interacción del usuario, también se realizará un seguimiento del uso de hashtags (García y Gómez, 2022).

El uso de las redes sociales en las elecciones de Colombia ha transformado la dinámica política del país. Antes el debate se centraba en los medios tradicionales, pero ahora se ha cambiado en gran medida al ámbito digital. La desconfianza hacia los medios tradicionales ha impulsado este cambio, con una creencia generalizada de que las redes sociales son más participativas y permiten una mayor influencia (Delgadillo, 2022). Sin embargo, esta transición no está exenta de dificultades. Simplificar los mensajes y priorizar el espectáculo sobre el contenido político real son preocupaciones importantes (Delgadillo, 2022). Además, existe una polarización de opiniones y la formación de filtros que limitan diferentes perspectivas (Delgadillo, 2022). A pesar de estos desafíos. Las redes sociales, junto con el cambio de los medios tradicionales a espacios en línea como Facebook y Twitter/X. Por ejemplo, cambio la situación local de los políticos como Rodolfo Hernández y William Dau, quienes utilizaron exitosamente las redes para conseguir apoyo y reconocimiento en las elecciones regionales (Botero, 2023). Sin embargo, para garantizar una política transparente y equitativa, también salen a la luz cuestiones como la desinformación y la polarización del discurso político, junto con la necesidad de una moderación eficiente y un debate libre en línea (Botero, 2023). Para decirlo de manera sencillo, las redes sociales han cambiado el panorama político en Colombia, presentando tanto beneficios como dificultades enormes.

Durante las campañas electorales de Madrid 2021, el uso de las redes sociales, en particular sitios como TikTok, en la política ha adquirido mayor importancia. Destacando la espectacularidad del contenido más que en problemas políticos, los partidos políticos están experimentando con estas plataformas de maneras novedosas (Fernández, 2022). Las ideas incluyen confrontar a los oponentes políticos, personalizar candidatos y usar hashtags. El enfrentamiento parece ser especialmente del agrado de los usuarios de TikTok, por el mayor nivel de participación (Fernández, 2022).

A diferencia de otras redes sociales como Twitter/X o Instagram, hay menos interacción directa entre las partes y sus seguidores. Se estudió el impacto del uso de Twitter/X por parte de los políticos españoles en la opinión pública durante las elecciones catalanas de diciembre de 2017. Se demostró que para aquellos con un impacto en las redes sociales, la calidad y relevancia de los temas discutidos importaban tanto como la cantidad de tweets. El principal problema que salió a la luz fue la independencia de Cataluña, que ocultó otros temas como el desempleo y la corrupción (Curiel y García, 2020). Si bien esta estrategia no siempre mostró las preocupaciones del público según lo informado por el CEI, los candidatos que dedicaron más tiempo a tuitear sobre este tema tuvieron el mayor impacto

en la opinión pública (Curiel y García, 2020). Este análisis destaca la importancia estratégica de las redes sociales en las campañas políticas para influir en la percepción pública y destaca la necesidad de relacionar los mensajes con las preocupaciones de la población.

El ejemplo de Chile sirve como evidencia el impacto de las redes sociales en la política. Twitter/X evolucionó hasta convertirse en una plataforma donde las personas denuncian casos de abusos, coordinaron acciones e intercambiaron información durante incidentes como la marea roja de Chiloé de 2016 (Allendes y Jorge, 2018) Si es difícil calcular el efecto de estas iniciativas en la movilización política, sí mostraron cómo se pueden utilizar las redes sociales para aumentar la conciencia pública y fortalecer las demandas de la sociedad civil. El análisis del uso de Twitter/X durante la marea roja de Chiloé mostró que la mayor parte de los tweets tenían como objetivo difundir información más que coordinar acciones políticas específicas (Allendes y Jorge, 2018) . Esto implica que, si bien las redes sociales pueden ser útiles para generar conciencia y opinión pública, si no se usan sabiamente, su potencial para la organización política puede verse limitado. Un estudio sobre el entorno electoral en Ecuador, el uso de redes sociales como Facebook se caracteriza por una baja interactividad y participación por parte de los ciudadanos, a pesar de que estas plataformas tienen la capacidad de ayudar a políticos y votantes a interactuar. Según el estudio del modelo de evaluación de Comunicación 2.0, la presencia, el crecimiento, la actividad o la participación de ninguna de las cuentas de Facebook de los solicitantes se encontraban en el rango óptimo (Benítez et al., 2022). Se observó que los relatos de los candidatos carecían de espacios de discusión y compromiso con los votantes y estaban destinados principalmente a difundir información. Aunque hubo diferencias en el uso de pronombres y ciertos temas, el lenguaje de la mayoría de los candidatos fue directo. Sin embargo, resultó que los candidatos prestaron poca atención a temas políticos específicos en sus posturas; en cambio, la mayoría de ellos se concentraron más en contenidos retóricos y emocionales (Benítez et al., 2022).

En el artículo “Cambridge Analytica y su impacto en las leyes de privacidad” de Timoteo Prezzavento explica cómo las redes sociales impactaron el discurso político y el alcance internacional del escándalo. Explica el papel de las redes sociales en la política, destaca cómo sitios como Facebook, Twitter/X y YouTube han hecho posible que los políticos se comuniquen directamente con el público mediante el envío de mensajes personalizados. Cambridge Analytica influyó en los resultados electorales en numerosos países mediante el uso de datos de Facebook para crear perfiles complejos y dirigir ciertos mensajes políticos (Prezzavento, s.f.). El asunto tuvo un impacto global significativo. Actualmente existe un debate global sobre la moralidad de utilizar datos personales con fines políticos debido a la creciente conciencia sobre la privacidad de los datos y la capacidad de las redes sociales para influir en la opinión pública. Además, como resultado, las normas y regulaciones de protección de datos (como el RGPD en la Unión Europea) han cambiado (Prezzavento, s.f.). Además de cambiar sus políticas de privacidad, las empresas de redes sociales han visto una disminución en la confianza del público como resultado de acusaciones de manipulación de datos.

La minería de datos es crucial para extraer datos de calidad a partir de conjuntos de datos masivos, es-

pecialmente en redes sociales como Twitter/X. Durante este proceso, se recopilan y procesan grandes volúmenes de datos generados por los usuarios, como tweets, encontrando patrones y tendencias (Blázquez, 2019). Sin embargo, en este estudio, el enfoque principal es la calidad de los datos extraídos, no su cantidad. La minería de datos se utiliza para la detección de tendencias en tiempo real y el análisis de los tweets (Blázquez, 2019), pero es fundamental asegurar que los datos sean precisos y fiables. Además, este estudio se enmarca dentro de los principios de ciencia abierta y reproducible. Implementamos metodologías transparentes y accesibles que permiten la verificación y replicación de los resultados. Este enfoque no solo mejora la calidad del análisis, sino que también promueve la colaboración y el avance del conocimiento en el campo del análisis de datos sociales y políticos.

## **Objetivos**

### **Objetivo general**

Evaluar la calidad de los microdatos recopilados de Twitter para los perfiles de los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la alcaldía de Bogotá.

### **Objetivos específicos**

1. Configurar un entorno de investigación reproducible para el análisis de datos, centrándose en mensajes (Tweets) relacionados con los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la alcaldía de Bogotá.
2. Caracterizar las herramientas para la extracción de datos “Twitter/X” de Gustavo Bolívar y Juan Daniel Oviedo durante la campaña electoral para la Alcaldía de Bogotá

## Marco teórico

La complicada dinámica de la explosión de información en Twitter/X durante una campaña política puede estudiarse desde diversos enfoques teóricos. Según la teoría de la comunicación política (Croteau & Hoynes, 2014), Twitter/X es un instrumento crítico de persuasión y apoyo debido a la naturaleza bidireccional de los intercambios entre los políticos y el público en general. Por otro lado, la teoría del establecimiento de agenda (McCombs y Shaw, 1972) enfatiza el papel crítico de los medios de comunicación, incluido Twitter/X, en la definición de cuestiones sociales y la creación de espacios para que los políticos definan agendas y discursos particulares. La teoría del framing (Entmann, 1993) enfatiza cómo la información proporcionada estratégicamente puede impactar la opinión pública, permitiendo a los políticos replantear los problemas y cambiar la percepción pública usando Twitter/X.

La gobernanza de la información o de los datos en Twitter/X (D'Agostino et al., 2017) es fundamental para una gestión eficaz de los datos y la seguridad de la información. Los medios en Twitter son fundamentales para difundir información y afectar la opinión pública (D'Agostino et al., 2017).

Finalmente, la apertura en la publicación de resultados y métodos es fundamental para la validez de la investigación, particularmente en entornos de Big Data . En este escenario, la narración de datos surge como una herramienta fundamental para transformar información complicada en narrativas comprensibles (Quijano, 2020).

## Enfoque metodológico

Este proyecto se guía por una metodología de diseño basada en investigación que se caracteriza por la creación de prototipos continuos. Este enfoque, en línea con el pensamiento de diseño y las epistemologías diseñistas, reconoce el diseño como un conocimiento en red, capaz de incorporar métodos cuantitativos, cualitativos, etnográficos e investigación acción participativa (Luna, 2019).

La investigación se desarrolla en ciclos de retroalimentación continua, lo que permite ajustar la metodología a medida que avanzamos en el estudio. Los prototipos, concebidos como laboratorios para explorar hipótesis, desempeñan un papel fundamental. No solo actúan como un medio para comunicar resultados, en consonancia con la propuesta de investigación desde el diseño, sino que también se utilizan como herramientas para habitar y explorar mundos posibles (Luna, 2019).

La metodología de diseño basada en investigación se estructura en varias fases

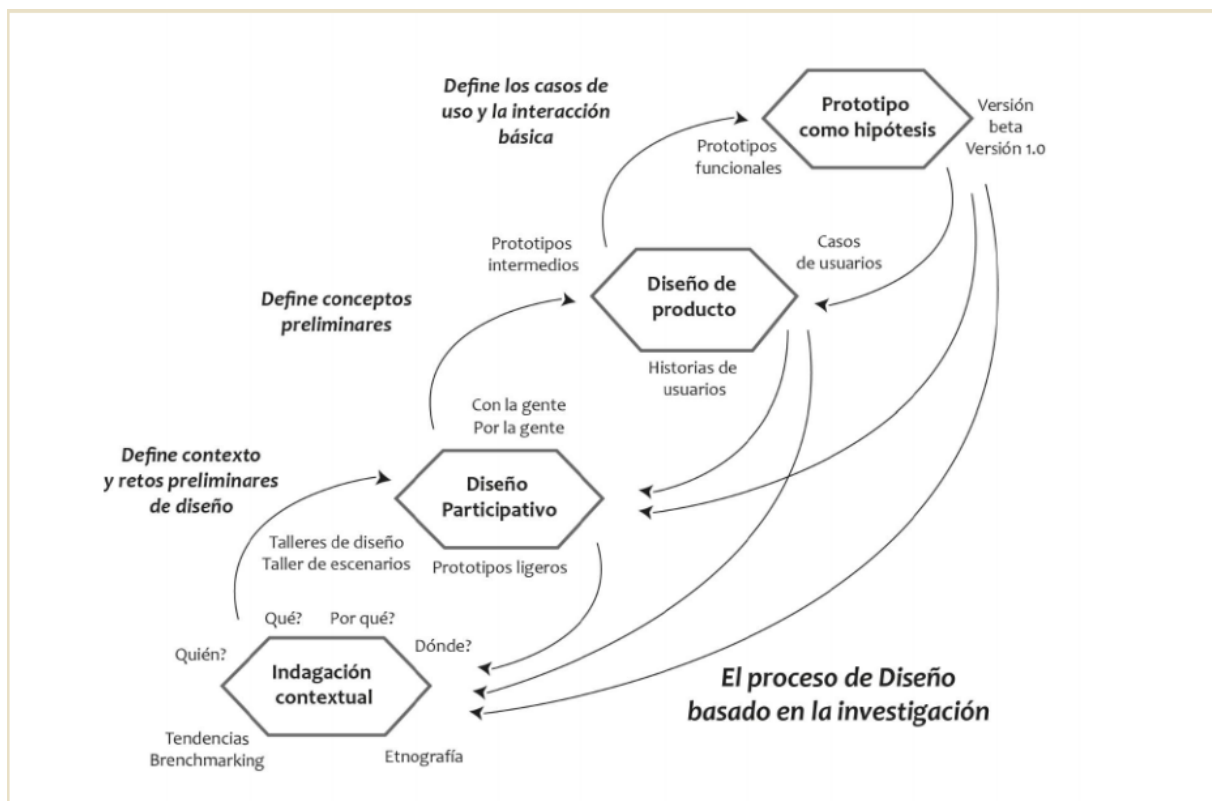


Figure 1: Imagen representativa de la metodología (Luna, 2019)

**Indagación contextual:** Esta fase implica la exploración del contexto sociocultural mediante técnicas etnográficas rápidas y conversaciones formales e informales (Luna, 2019). Utilizando herramientas de minería de datos para examinar grandes volúmenes de datos creados por usuarios de la red social, se



llevó a cabo un análisis de calidad de los tweets de Twitter/X durante la campaña electoral.

**Diseño participativo:** Aquí, se fomenta la colaboración con los interesados a través de talleres y la creación de prototipos ligeros o mentales. El objetivo es formular hipótesis sobre posibles soluciones, incorporando la perspectiva de quienes están directamente involucrados (Luna, 2019). Utilizamos técnicas de minería de datos para enfocar en la calidad y la organización de los datos textuales creados de Twitter/X durante las candidaturas a la alcaldía de Bogotá. En esta fase, nos centramos directamente en el diseño del producto sin someter nuestros prototipos a la evaluación de una comunidad o pluralidad de personas.

**Diseño de producto:** Durante esta etapa, se lleva a cabo la creación de prototipos tempranos por parte del equipo del proyecto. Aunque se mantenga cierta distancia de los interesados, se fomentan discusiones técnicas especializadas para asegurar la viabilidad del diseño Luna, 2019, citando a Leinonen (2008). Se crea un entorno reproducible con el fin de almacenar, analizar y organizar los datos encontrados en los perfiles.

**Prototipo como hipótesis:** En esta fase, los prototipos se presentan a los interesados para su validación. Se destaca la naturaleza hipotética de los prototipos, lo que permite evaluaciones continuas dentro de los límites del proyecto Luna, 2019, citando a Leinonen (2008). Se iteró en los prototipos, mirando qué tanto ellos podían efectivamente dar cuenta de la calidad de los datos y qué elemento algorítmicos completaban la narrativa o donde estaban las limitaciones, bien en los conocimientos para mejorar los análisis o bien en las herramientas para dicho análisis.

Este enfoque metodológico se adapta a las necesidades específicas de la investigación sobre la calidad de los microdatos extraídos de Twitter/X para los perfiles de los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la Alcaldía de Bogotá del año 2023, permitiendo una exploración profunda y una validación constante de los hallazgos.

Estas modificaciones demuestran la aplicación única de las fases de la metodología de diseño general del dominio público al estudio de análisis de minería de datos sobre la influencia política en Twitter/X durante la contienda por la Alcaldía de Bogotá.

## Desplegando un entorno de investigación reproducible

El entorno de investigación reproducible para esta tesis está compuesto de los siguientes elementos:

1. **Scoop:** Gestor de instalación de software libre, de código abierto y/o libre distribución.
2. **Fossil y ChiselApp:** Sistema de control de versiones y publicación de borradores de la tesis y sus narrativas de datos
3. **Glamorous Toolkit y Pharo:** Entorno de desarrollo moldeable con herramientas visuales e interactivas para programación y gestión del conocimiento.
4. **Grafoscopio:** Herramienta que incluye ExoRepo y MiniDocs para la organización y visualización de documentos.
  - **ExoRepo:** Utilidad que facilita la instalación de repositorios hospedados en infraestructuras de código autónomas.
  - **MiniDocs:** Sistema diseñado para mejorar las capacidades de documentación de proyectos, integrado en Grafoscopio.
5. **WindTerm:** Aplicación para la gestión de terminales y líneas de comandos.
6. **Zettlr:** Software para la escritura y organización de textos académicos y notas.
7. **Pandoc:** Conversor de documentos entre diferentes formatos, como Markdown, HTML y PDF.
8. **Tectonic:** Motor de composición tipográfica para la creación de documentos LaTeX.
9. **Eisvogel Template:** Plantilla para la generación de documentos PDF con un diseño elegante.
10. **LiteXL:** Editor de texto ligero y extensible para la edición de archivos.
11. **Hypothesis:** Herramienta para la creación comentarios en investigaciones.
12. **Flameshot:** Herramienta para tomar pantallazos
13. **Publicando reproducible:** Como exportamos el documento desde Markdown a HTML y PDF y cómo se hacen con Fossil.

A continuación, se detallará exhaustivamente cada una de las herramientas mencionadas, incluyendo cómo fueron instaladas y utilizadas durante el desarrollo de la tesis. Esto permitirá comprender en profundidad el papel de cada herramienta en el proceso de investigación reproducible.

### Scoop

Scoop es una aplicación de línea de comandos para Windows que facilita la instalación de software al eliminar las ventanas emergentes de permisos y los tutoriales de instalación. Automatiza la instalación de dependencias, realiza las configuraciones necesarias y puede crear configuraciones recurrentes para personalizar el entorno. Es especialmente útil para instalar aplicaciones “portátiles”, pero también funciona con instaladores estándar.

Para su instalación, se puede realizar de dos maneras. La primera es directamente desde la página principal, mientras que la otra se lleva a cabo mediante el siguiente comando:

```
iwr -useb get.scoop.sh | iex
```

Scoop se utilizó para instalar todos los software para el entorno reproducible, un gran ejemplo es Pandoc:

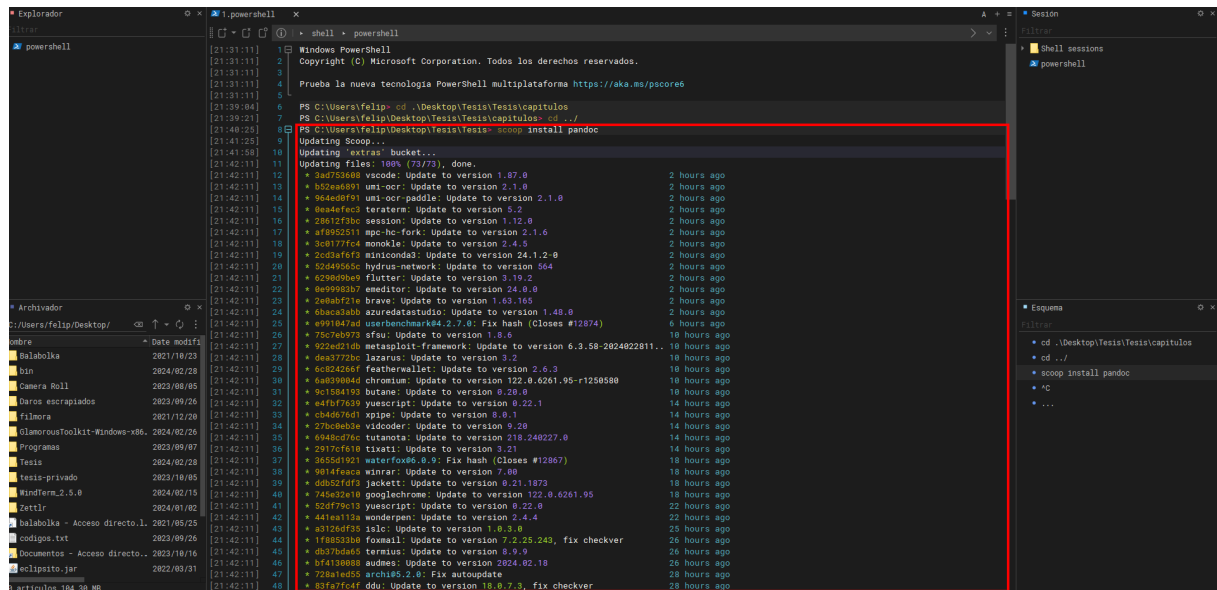
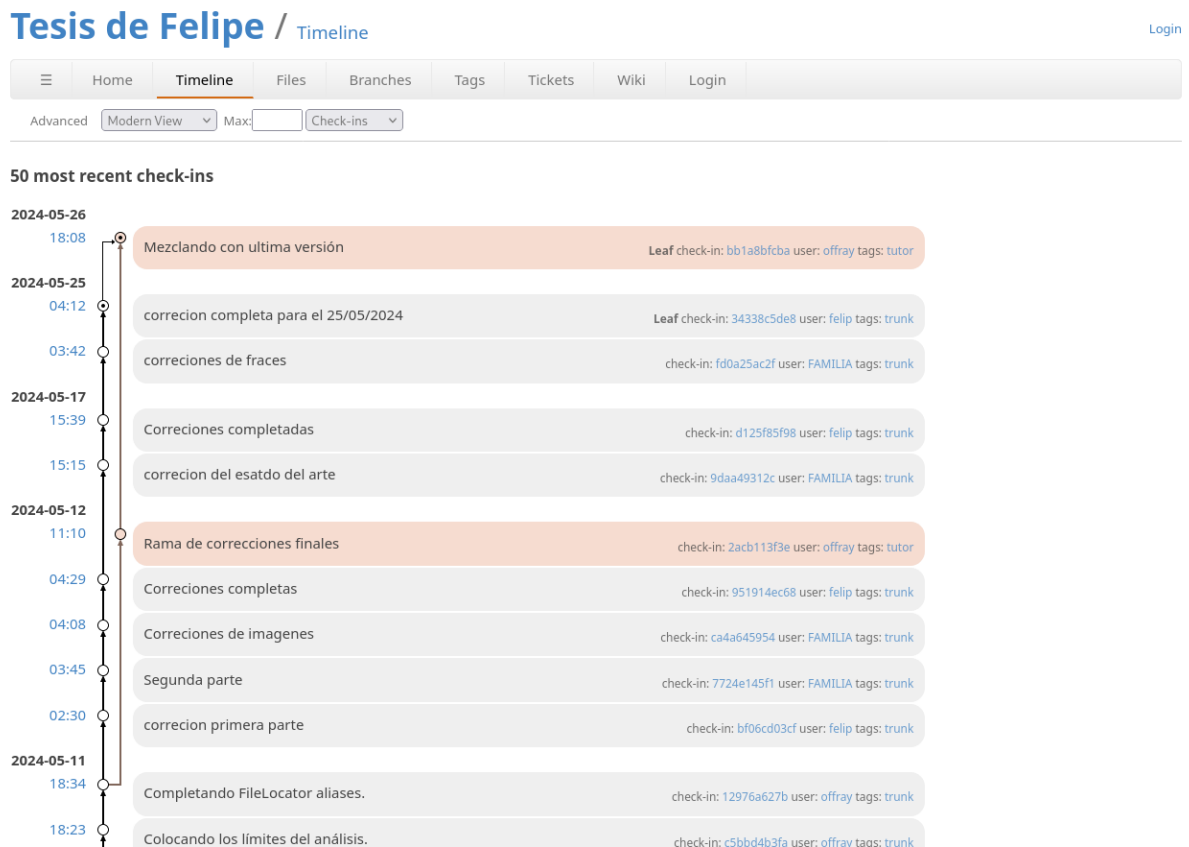


Figure 2: Figura 1: Uso de Scoop

## Fossil y ChiselApp

Fossil es un sistema de control de versiones (SCM) que rastrea los cambios, el progreso y el desarrollo del proyecto. Destaca por su alta calidad y es muy reconocido por su uso en ChiselApp, una plataforma de alojamiento y lanzamiento de proyectos. Aunque el Fossil SCM se considera un dispositivo de alta calidad, no recibe el crédito que merece en el sector tecnológico.

ChiselApp es una plataforma de alojamiento y lanzamiento de proyectos que ejecuta Fossil SCM como sistema de control de versiones. Permite a los usuarios realizar un seguimiento de los cambios en los archivos, comunicar el progreso y la evolución del trabajo académico y guardar los datos recopilados mediante herramientas de scraping. ChiselApp permite crear repositorios públicos y privados, lo que facilita el mantenimiento y la colaboración en proyectos académicos. En la imagen que se presentará a continuación se muestra el historial de cambios subidos a ChiselApp por los autores y el tutor del proyecto de grado. En ella, se pueden evidenciar las modificaciones de archivos y sus ramificaciones.



**Figure 3:** Figura 1: Uso de ChiselApp

Para la instalación y sincronización de Fossil se usó los siguientes comandos:

```
scoop install main/fossil
```

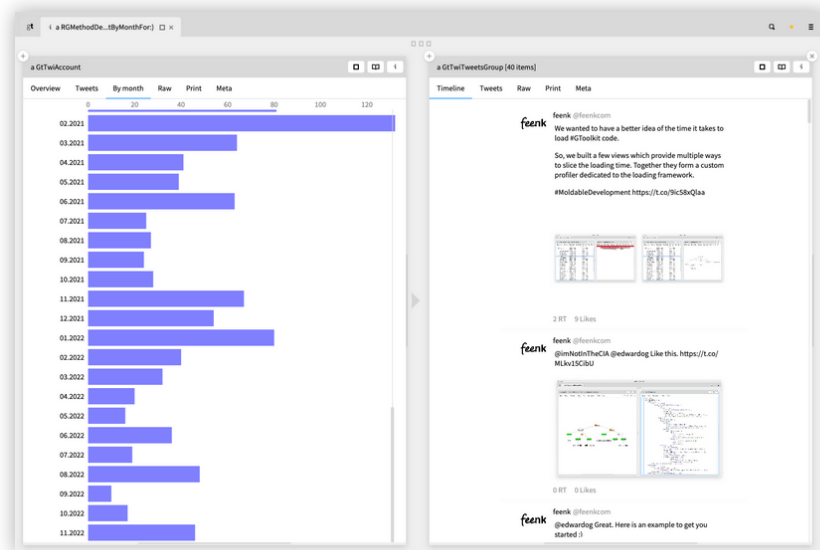
## Glamorous Toolkit y Pharo

Glamorous Toolkit (GT) es un entorno de desarrollo moldeable que ofrece herramientas visuales e interactivas que se pueden utilizar para abordar una variedad de preguntas, incluida la programación, la extracción de datos y la gestión del conocimiento. GT pretende hacer comprensibles los sistemas mediante experimentos adaptados a cada situación. Se implementa utilizando Pharo, un lenguaje de programación dinámico orientado a objetos. Pharo ofrece un entorno interactivo y un sistema de desarrollo integrado, lo que facilita el diseño de aplicaciones complejas. GT y Pharo trabajan juntos para brindar una experiencia de desarrollo única y altamente personalizada. Pharo se utilizó para realizar análisis de datos específicos. La integración de Glamorous Toolkit (GT) con Pharo permitió una

experiencia de desarrollo única y altamente personalizada, optimizando la creación y manipulación de datos para la investigación.

## Data exploration

Data holds valuable patterns that can only be unearthed when they are made visible. Visualization is important but it becomes a superpower only when you combine it with modeling and exploration.



Exploring the data exported from a Twitter account.

**Figure 4:** Figura 1: Uso de Glamorous Toolkit y Pharo

## Grafoscopio

Ahora se instalará Grafoscopio, Grafoscopio es una herramienta interactiva flexible de visualización de datos y documentación.

# Grafoscopio

publicación e investigación reproducible ⊕ documentación interactiva ⊕ visualización flexible y ágil de datos

### Escribe, estructura y unifica



Distintos tipos documentos todos ellos en un sólo sitio. Jeraquiza, agrupa y visualiza aquello que quieres y crea diferentes vistas de la misma información.

[MIRA UN EJEMPLO >](#)

### Cálcula y visualiza



Puedes conectar varias fuentes de información heterogeneas e integrarlas mediante visualizaciones poderosas y flexibles.

[MIRA UN EJEMPLO >](#)

### Colabora y publica



Las plantillas y el control histórico de cambios, facilitarán el trabajo individual y colectivo, así como su publicación en PDF y HTML.

[MIRA UN EJEMPLO >](#)

### Explora y modifica



Usa el poderoso entorno interactivo de programación integrado, para ver cómo está hecha la herramienta y adaptarla a diversas necesidades.

[MÁS INFO >](#)

Grafoscopio es una herramienta amoldable para documentación interactiva y visualización de datos, que está siendo usada para ciencia abierta, ciudadanas y de garage, investigación reproducibles, (h)ac(k)tivismo, innovación abierta y comunitaria,

Se utiliza en una variedad de disciplinas, incluida la ciencia abierta, la investigación reproducible. ExoRepo es una utilidad que facilita la instalación de repositorios hospedados en infraestructuras de código autónomas.

Pero antes de eso debemos instalar Exorepo, esto se debe instalar en Lepiter de Glamorous Toolkit

```
location := FileLocator localDirectory .
(IceRepositoryCreator new
  location: location;
  remote: (IceGitRemote url: 'https://code.sustrato.red/Offray/ExoRepo.git');
  createRepository)
register.
```

```
Metacello new
  repository: 'gitlocal://', location fullName;
  baseline: 'ExoRepo';
  load
```

Ahora se instala MiniDocs, un sistema diseñado para mejorar las capacidades de documentación de

Lepiter.

ExoRepo install: 'MiniDocs'.

## WindTerm

WindTerm es un programa de terminal diseñado para un rendimiento rápido y eficiente en Windows y WSL (Subsistema de Windows para Linux). Destaca por su capacidad para procesar comandos rápidamente y utilizar una memoria mínima, lo que lo hace perfecto para tareas que necesitan grandes cantidades de texto. WindTerm proporciona transferencias de archivos locales y SFTP rápidas, lo que lo convierte en una opción eficiente para la administración de archivos. También tiene la capacidad de autocompletar credenciales para acelerar el proceso de inicio de sesión, aunque se recomienda su uso en entornos seguros debido a riesgos potenciales de seguridad.

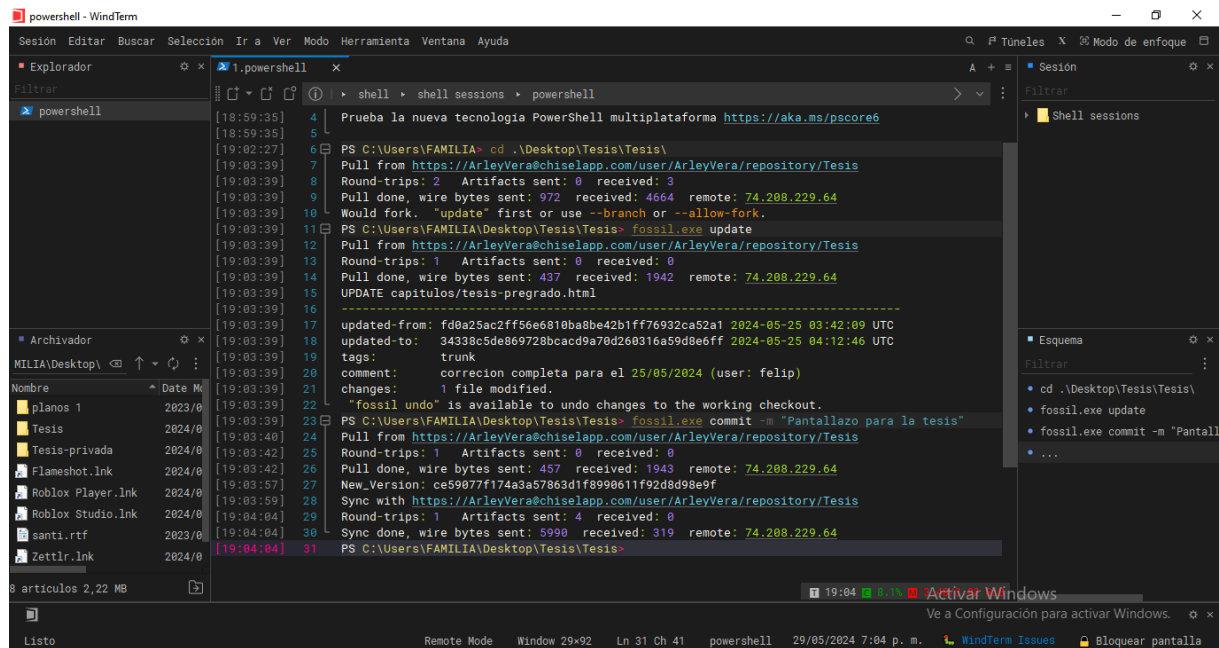


Figure 5: Figura 1: WindTerm

## Zettlr

Zettlr es una herramienta de escritura todo en uno que le ayuda a crear publicaciones que incluyen publicaciones de blogs, artículos de revistas, artículos académicos, tesis y libros completos. Incluye gestión de proyectos de redacción, la mejor asistencia para citas, privacidad de datos, conexión con administradores de referencias, exportación flexible y una interfaz de usuario diseñada para brindar

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

eficiencia, velocidad y productividad. Zettlr es un software gratuito y de código abierto que mantiene sus notas en su computadora sin sincronización forzada en la nube ni tarifas ocultas. Además, protege su privacidad al no rastrear ni experimentar con sus datos. La herramienta se puede adquirir por la pagina oficial de Zettlr. En la imagen a continuación se observa la interfaz de Zettlr, donde se muestran los archivos Markdown abiertos del trabajo de grado, con la posibilidad de modificarlos.

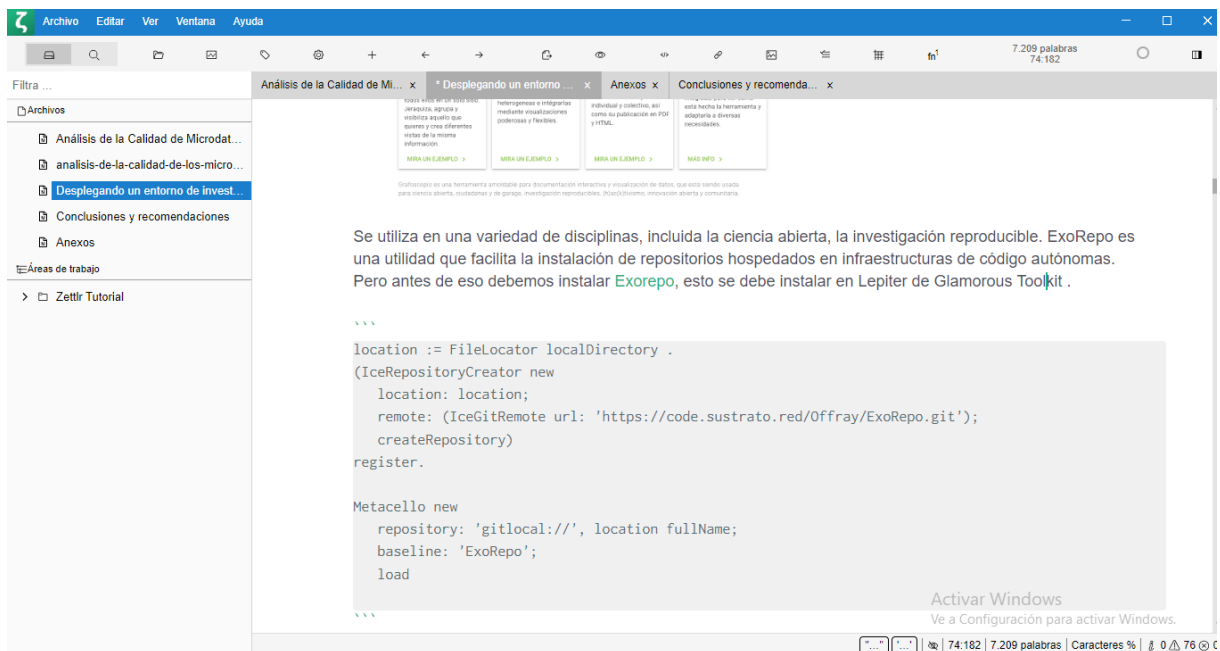


Figure 6: Figura 1: Zettlr

## Pandoc

Pandoc es una aplicación de línea de comandos y una biblioteca Haskell que puede convertir documentos entre varios formatos procesamiento de textos, incluidos Markdown, HTML, LaTeX y Word docx. Pandoc es extremadamente flexible y modular, por lo que los usuarios pueden agregar rápidamente formatos de entrada y salida adicionales. También puede crear documentos separados o fragmentos y generar archivos PDF desde LaTeX, ConTeXt, roff ms o HTML como formatos intermedios.

Para su instalación, se puede realizar de dos maneras. La primera es directamente desde la página principal, mientras que la otra, se lleva a cabo mediante el siguiente comando:

```
scoop install pandoc
```



# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

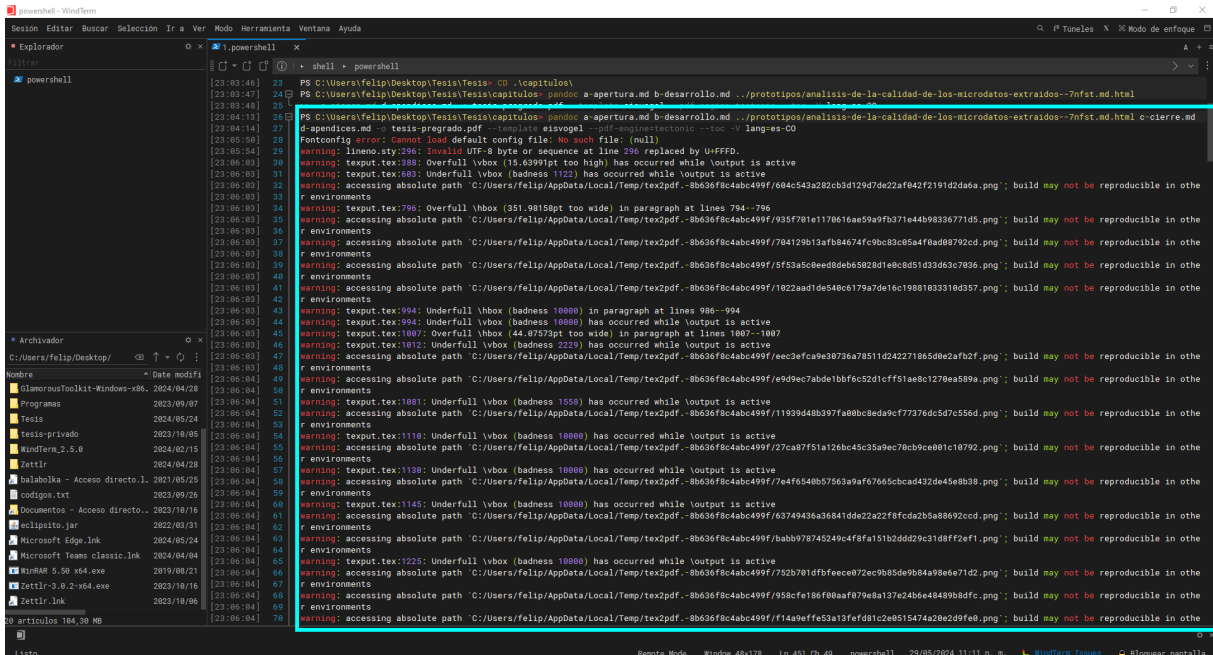


Figure 7: Figura 1: Scoop

## Tectonic

Un motor de composición de documentos basado en TeX/LaTeX llamado Tectonic facilita la creación de textos que sean rigurosamente científicos y tipográficamente precisos. La carga automática de archivos de soporte, la producción de documentos reproducibles, la lógica de creación inteligente, la integración de GitHub Actions y la distribución de licencias MIT son algunas de las características importantes.

Para su instalación se lleva a cabo mediante el siguiente comando:

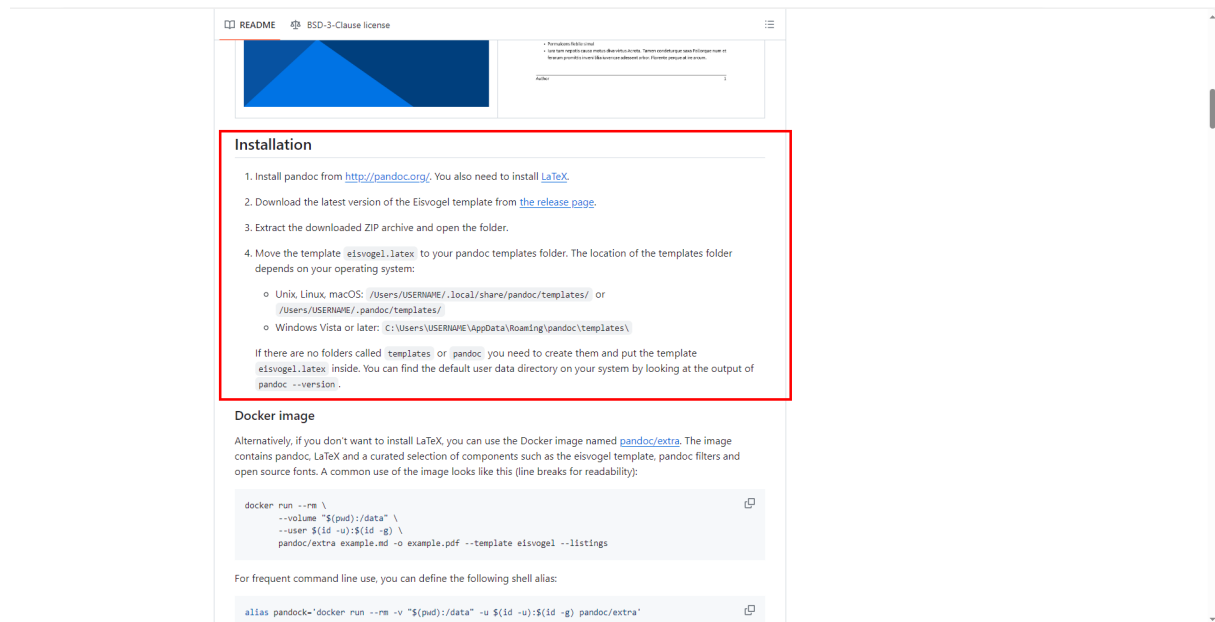
```
scoop install main/tectonic
```

## Eisvogel Template

La plantilla Eisvogel LaTeX está diseñada para utilizar Pandoc para convertir archivos Markdown a PDF o LaTeX. Se concentra especialmente en la redacción de ejercicios y apuntes del curso, con especial atención a la informática. La plantilla tiene una serie de características, como soporte para resaltado de sintaxis en bloques de código, incluyen imágenes de fondo y personalización de la página de título. También brinda oportunidades para manipular la apariencia del documento final a través de

variables personalizadas. Junto con Pandoc y LaTeX, la plantilla Eivogel se puede descargar e instalar. También se puede acceder a él a través de una imagen de Docker que contiene Pandoc, LaTeX y otros archivos necesarios.

Para su instalación se lleva a cabo mediante de su paginá oficial de descargar y siguiendo sus parámetros para su instalación.



**Figure 8:** Figura 1: Eivogel Template

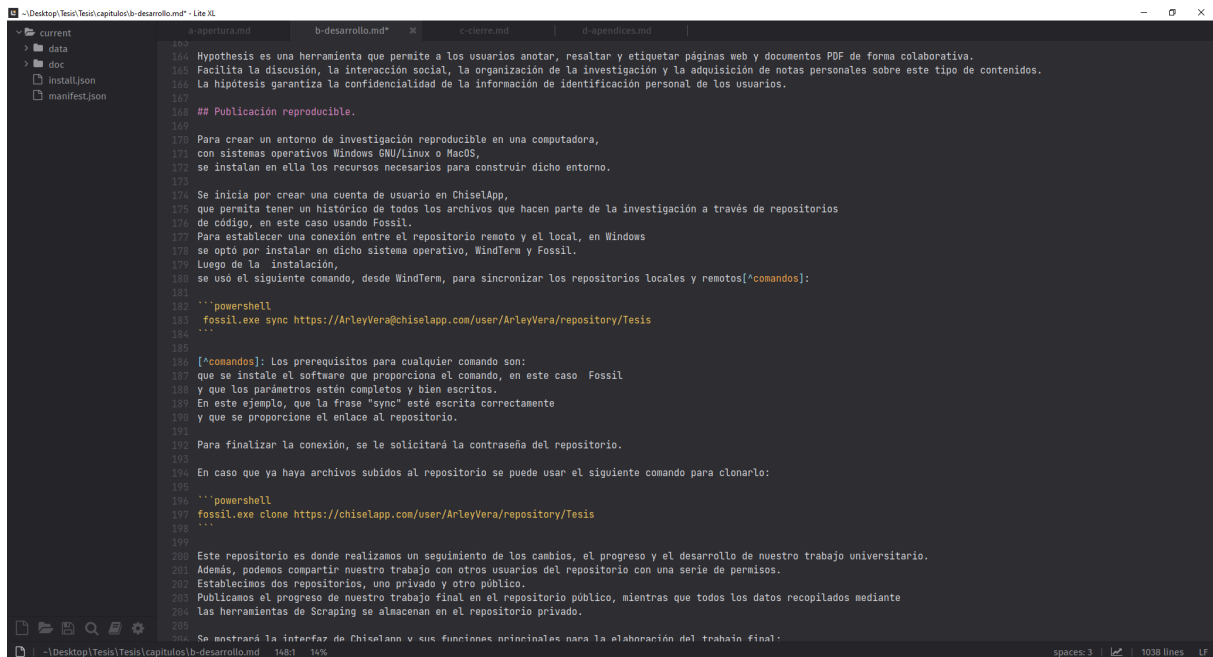
## LiteXL

Lite XL es un editor de texto liviano construido principalmente en Lua y diseñado para ser funcional, estéticamente agradable, pequeño y rápido. Permite la personalización mediante complementos y temas de color, que pueden cambiarse fácilmente modificando el módulo de usuario. Para su instalación se necesita ejecutar el siguiente comando:

```
scoop install lite-xl
```

Interfaz de LiteXL:

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



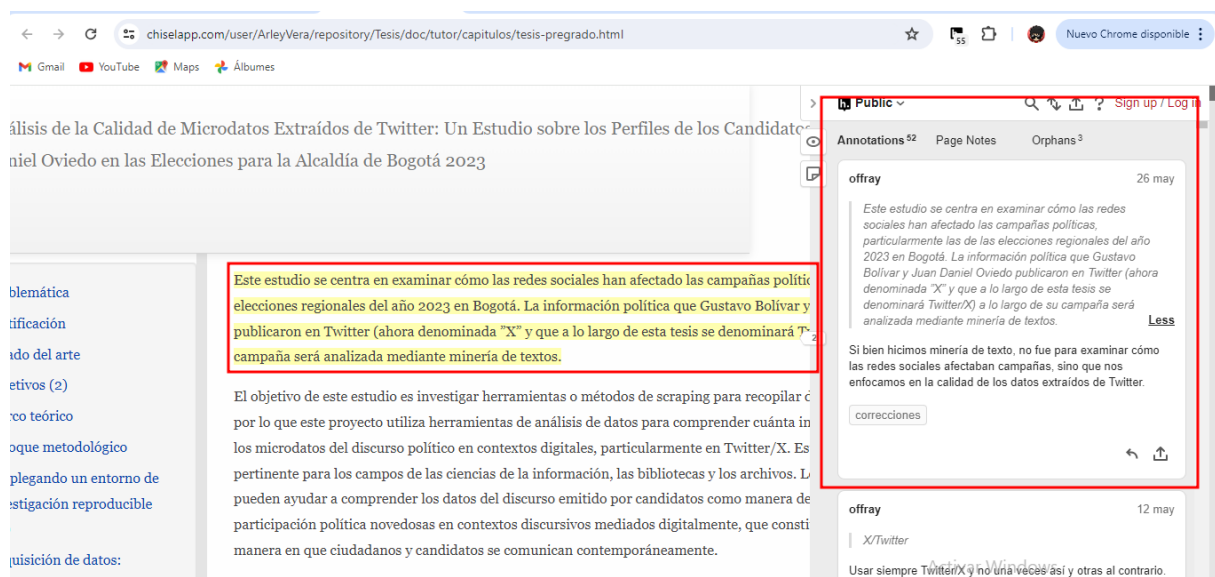
```
164 Hypothesis es una herramienta que permite a los usuarios anotar, resaltar y etiquetar páginas web y documentos PDF de forma colaborativa.
165 Facilita la discusión, la interacción social, la organización de la investigación y la adquisición de notas personales sobre este tipo de contenidos.
166 La hipótesis garantiza la confidencialidad de la información de identificación personal de los usuarios.
167
168 ## Publicación reproducible.
169
170 Para crear un entorno de investigación reproducible en una computadora,
171 con sistemas operativos Windows GNU/Linux o MacOS,
172 se instalan en ella los recursos necesarios para construir dicho entorno.
173
174 Se inicia por crear una cuenta de usuario en ChiselApp,
175 que permita tener un histórico de todos los archivos que hacen parte de la investigación a través de repositorios
176 de código, en este caso usando Fossil.
177 Para establecer una conexión entre el repositorio remoto y el local, en Windows
178 se optó por instalar en dicho sistema operativo, WindTerm y Fossil.
179 Luego de la instalación,
180 se usó el siguiente comando, desde WindTerm, para sincronizar los repositorios locales y remotos[*comandos]:
181
182 ```powershell
183 fossil.exe sync https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
184 ```
185
186 [*comandos]: Los prerequisites para cualquier comando son:
187 que se instale el software que proporciona el comando, en este caso Fossil
188 y que los parámetros estén completos y bien escritos.
189 En este ejemplo, que la frase "sync" esté escrita correctamente
190 y que se proporcione el enlace al repositorio.
191
192 Para finalizar la conexión, se le solicitará la contraseña del repositorio.
193
194 En caso que ya haya archivos subidos al repositorio se puede usar el siguiente comando para clonarlo:
195
196 ```powershell
197 fossil.exe clone https://chiselapp.com/user/ArleyVera/repository/Tesis
198 ```
199
200 Este repositorio es donde realizamos un seguimiento de los cambios, el progreso y el desarrollo de nuestro trabajo universitario.
201 Además, podemos compartir nuestro trabajo con otros usuarios del repositorio con una serie de permisos.
202 Establecimos dos repositorios, uno privado y otro público.
203 Publicamos el progreso de nuestro trabajo final en el repositorio público, mientras que todos los datos recopilados mediante
204 las herramientas de Scraping se almacenan en el repositorio privado.
205
206 Se mostrará la interfaz de ChiselApp y sus funciones principales para la elaboración del trabajo final.
```

Figure 9: Figura 1: LiteXL

## Hypothesis

Hypothesis es una herramienta que permite a los usuarios anotar, resaltar y etiquetar páginas web y documentos PDF de forma colaborativa. Facilita la discusión, la interacción social, la organización de la investigación y la adquisición de notas personales sobre este tipo de contenidos. La hipótesis garantiza la confidencialidad de la información de identificación personal de los usuarios.

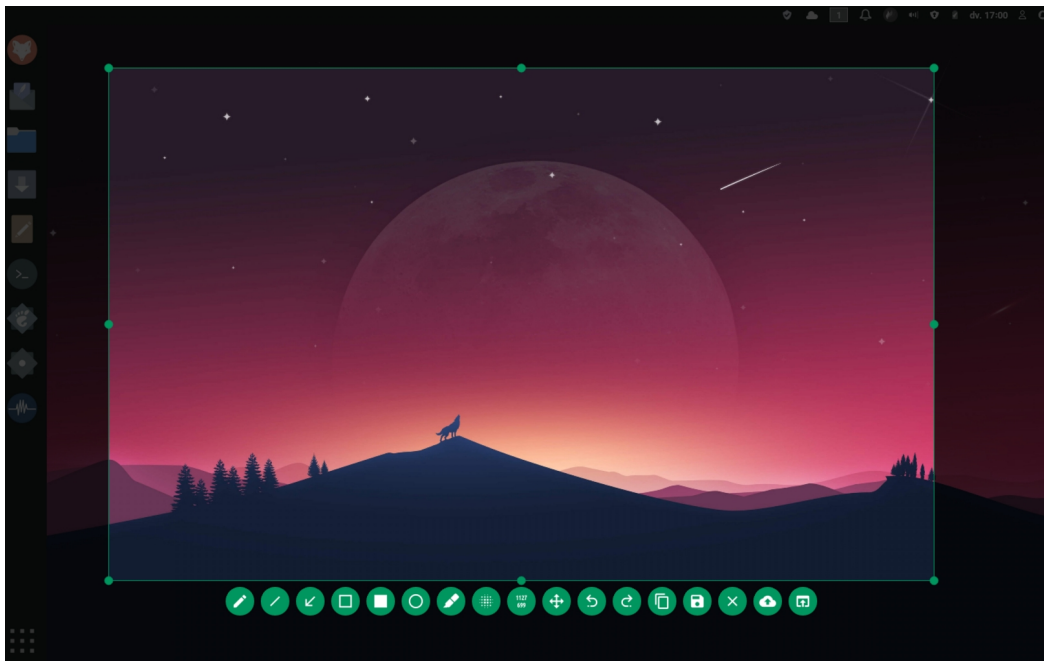
# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 10:** Hypothesis, siendo empleada para los comentarios y corrección por parte del tutor y la realimentación al estudiante.

## Flameshot

Flameshot es una herramienta de captura de pantalla altamente personalizable y fácil de usar. Ofrece numerosas opciones de personalización, como la interfaz de color, la selección de botones, atajos de teclado y la forma en que se guardan las imágenes. Su configuración es accesible y permite modificar estos aspectos según las necesidades del usuario. En la imagen a continuación se puede evidenciar el uso de una herramienta para tomar un pantallazo como ejemplo. Se utilizó esta herramienta para todas las imágenes o capturas de pantalla encontradas en el trabajo de grado.



**Figure 11:** Figura 1: Interfaz de Flameshot I

### **Publicación reproducible.**

Si bien hemos mostrado, en las secciones previas las herramientas individuales y sus usos en esta investigación. Este apartado se encarga de mostrar los flujos de trabajo que conectaron las herramientas.

Se inicia por crear una cuenta de usuario en ChiselApp, que permita tener un histórico de todos los archivos que hacen parte de la investigación a través de repositorios de código, en este caso usando Fossil. Para establecer una conexión entre el repositorio remoto y el local, en Windows se optó por instalar en dicho sistema operativo, WindTerm y Fossil. Luego de la instalación, se usó el siguiente comando, desde WindTerm, para sincronizar los repositorios locales y remotos:

```
fossil.exe sync  
↪ https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
```

Los requisitos para cualquier comando son: que se instale el software que proporciona el comando, en este caso Fossil y que los parámetros estén completos y bien escritos. En este ejemplo, que la frase “sync” esté escrita correctamente y que se proporcione el enlace al repositorio.

Para finalizar la conexión, se le solicitará la contraseña del repositorio.

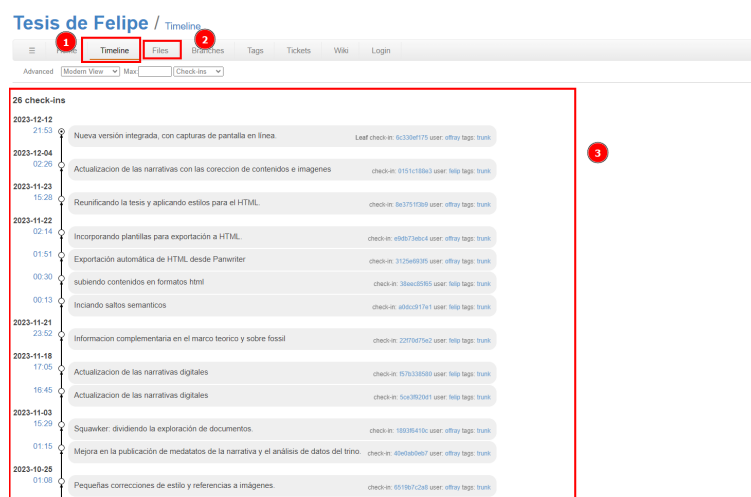
En caso de que ya haya archivos subidos al repositorio se puede usar el siguiente comando para clonarlo:

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

fossil.exe clone https://chiselapp.com/user/ArleyVera/repository/Tesis

Este repositorio se hace seguimiento de los cambios, el progreso y el desarrollo del trabajo universitario. Además, se puede compartir el trabajo con otros usuarios del repositorio con una serie de permisos. Se estableció dos repositorios, uno privado y otro público. Publicando el progreso del trabajo final en el repositorio público, mientras que todos los datos recopilados mediante de las herramientas de Scraping se almacenan en el repositorio privado.

Se mostrará la interfaz de Chiselapp y sus funciones principales para la elaboración del trabajo final:

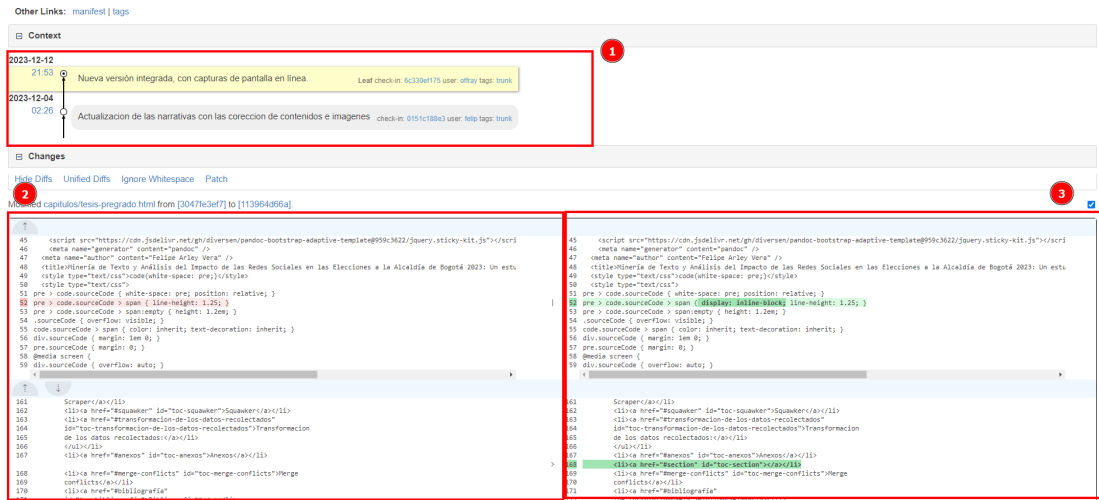


**Figure 12:** Figura 20: Repositorio publico

1. La “Línea de tiempo” muestra los cambios realizados en los archivos.
2. Todas las carpetas y archivos asociados con el repositorio se pueden encontrar en la sección “Files”.
3. Esta sección muestra la línea de tiempo de las modificaciones de archivos.

Al hacer clic en el enlace junto a la opción “Check in”, aparecerá una nueva interfaz con todos los datos e información actualizados. Además, se podrá comparar visualmente dónde estará la versión actualizada del mismo archivo con su versión anterior para ver los cambios realizados.

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 13:** Repositorio público con la diferencia entre dos versiones sucesivas de la tesis.

1. Línea del tiempo de archivo seleccionado.
2. Versión anterior de archivo y resaltando cual característica fue cambiada.
3. Versión de archivo actualizando resaltando la característica cambiada.

Es fundamental señalar que el uso de repositorios públicos y privados nos proporciona un enfoque integral para la gestión de datos y trabajo, que es necesario para garantizar la transparencia y reproducibilidad del estudio. Además, al publicar el trabajo en un repositorio público, contribuimos al intercambio de conocimiento y facilitamos la colaboración con otros académicos interesados en el campo de estudio. Para fusionar prosa con historias de datos, el documento de Google Docs se transformó a Markdown. Markdown es un lenguaje de marcado simple y comprensible que permite formatear texto de manera liviana. No sólo facilita la presentación de la información, sino que también ayuda a mejorar estéticamente el contenido del trabajo de grado. La conversión a Markdown permite integrar códigos de prueba para la recopilación y recuperación de datos mediante código o comandos. Esto es fundamental para la investigación, ya que permite automatizar operaciones y preservar un registro organizado de los procesos. Al mismo tiempo, Markdown hace que los datos y los resultados sean más atractivos y sencillos, lo que facilita la comprensión tanto para los académicos como para las personas interesadas en el trabajo.

Se utilizó Zettlr para dividir el trabajo en tres partes después de convertir el contenido del trabajo de grado a Markdown. Zettlr es un editor de texto gratuito y de código abierto que admite el formato Markdown y se adapta a varios sistemas de productividad como PARA, GTD y Zettelkasten. Permite el desarrollo de archivos de texto interconectados con enlaces internos y externos, formatos, tablas, gráficos, HTML y otras características sin necesidad de un proveedor. Zettlr le permite exportar fácilmente a muchos formatos, como PDF y HTML, así como ingresar caracteres especiales y texto automático. Puede utilizarse para ejecutar enfoques de productividad como Zettelkasten, PARA y GTD.

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

Sin embargo, tiene ciertas limitaciones, como la falta de conexiones a puntos específicos de los documentos, la falta de una versión móvil, la falta de transclusión, la falta de un lector gráfico y la imposibilidad de generar HTML.

La interfaz de Zettlr continúa enfatizando el formato Markdown, así como la funcionalidad principal de la herramienta, que incluye guardar, abrir y modificar archivos.

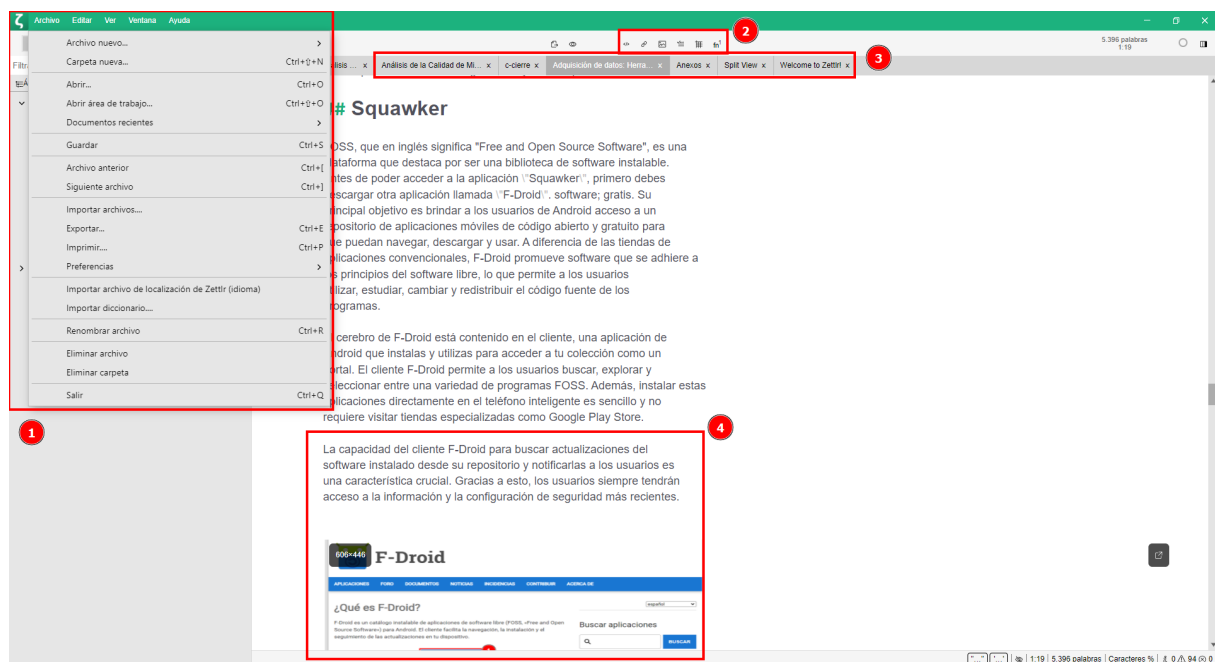


Figure 14: Figura 1: Zet

Interfaz Zettlr:

1. Opciones guardar, exportar, importar y abrir archivos
2. Herramientas como insertar imágenes por medio de un comando
3. Se te permite abrir varios archivos y modificarlos a la vez
4. Se permite modificar los archivos directamente

El trabajo se divide en tres partes:

**“a-apertura”** : Contiene la sección metodológica completa del trabajo de pregrado, incluyendo la introducción, estado del arte, objetivos y enfoque metodológico.

**b-desarrollo”**: Esta sección contiene información sobre herramientas de raspado y comparaciones. También se incluye una evaluación de la calidad de los datos del scrap y su correspondiente narrativa de datos. También menciona las aplicaciones y herramientas que se emplearon para lograr los resultados deseados.



**“c-cierre”:** Esta sección cubre los resultados, conclusiones y bibliografía del trabajo de pregrado.

Luego de la separación de los contenidos, se realizó un examen exhaustivo para encontrar fallas, como problemas de escritura y espacios adicionales, y organizar los contenidos de manera significativa. Luego, las tres partes se combinaron en un archivo PDF para poder visualizar el resultado y el contenido del trabajo de grado.

Luego se utilizó WindTerm para transferir los archivos al repositorio público, facilitando y agilizando dicho proceso. WindTerm es un software de terminal que se ejecuta de forma rápida y eficiente en entornos Windows y WSL (Subsistema de Windows para Linux). Se distingue por su capacidad para ejecutar comandos rápidamente y requerir un mínimo de memoria, lo que lo hace perfecto para trabajos que involucran grandes cantidades de texto.

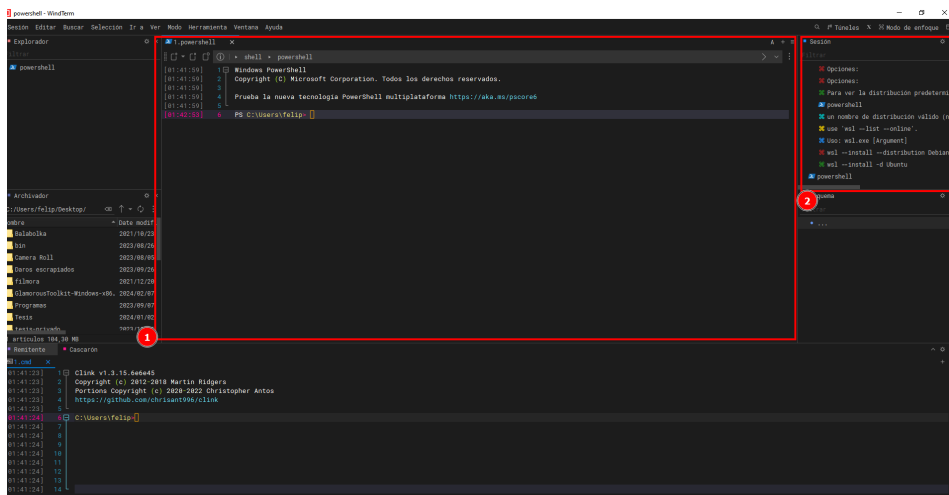


Figure 15: Figura 1: WindTerm

1. Apartado para escribir los comandos.
2. Apartado para seleccionar el tipo de terminal necesario.

Luego se puede utilizar LittleXL para acelerar el proceso, ya que Zettlr puede resultar pesado en algunos dispositivos. Si experimenta ralentizaciones u otras preocupaciones especiales, utilice LittleXL, que es una herramienta mucho más ligera y fácil de usar.

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

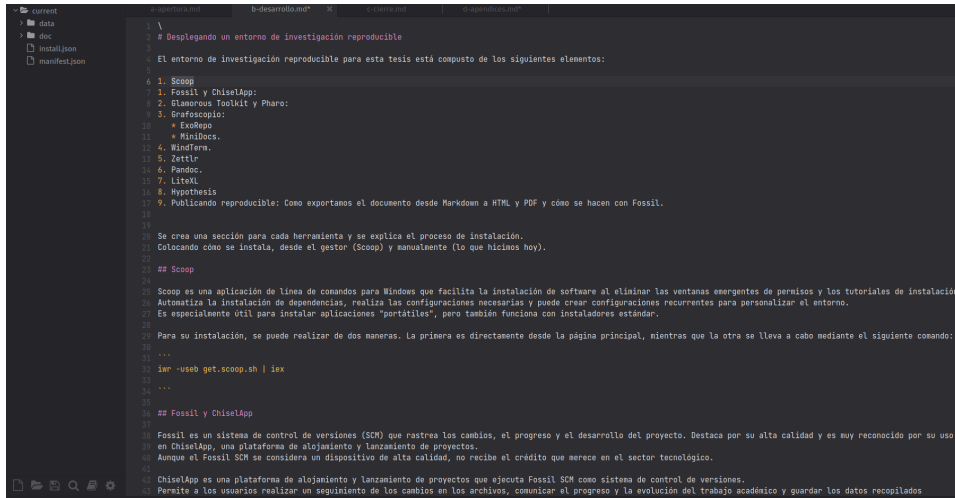


Figure 16: Figura 1: Interfaz de LittleXL

Luego se descargó el Glamorous Toolkit y se utilizó para evaluar los datos recuperados. Sin embargo, para lograr este objetivo, es esencial instalar las siguientes herramientas adicionales:

- Scoop.
- Grafoscopio.
- ExoRepo.
- MiniDocs.

Una vez instalados estos programas, puede comenzar a analizar los datos recuperados. Una vez que se complete el escaneo, siga los pasos a continuación para obtener la página que contiene la información o las instrucciones utilizadas para el escaneo.

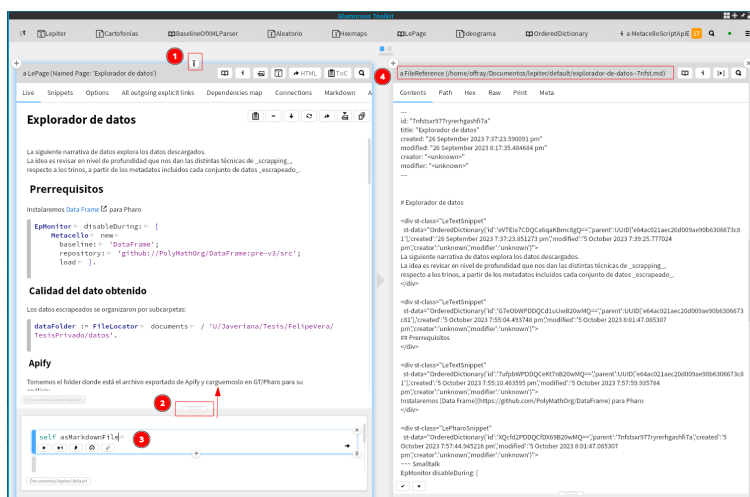


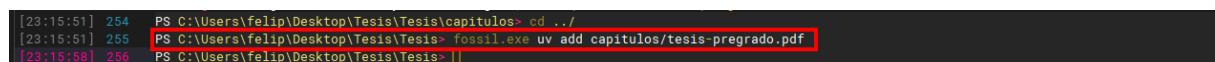
Figure 17: Figura 1: Glamorous Toolkit

1. Clickear el botón de inspector de la página que está en la mitad de la misma.
2. Arrastrar en handler desde el borde inferior de la página hacia arriba de modo que aparezca el playground de la página que está en 3.
3. Escribir en el playground de la página “self asMarkdownFile”. Eso producirá el archivo que está en 4, dentro de la subcarpeta Documentos/lepiter
4. Copiar el archivo que aparece en 4 a tu repositorio en la subcarpeta prototipos.

Después de tener todos los archivos con su respectiva información correspondiente se debe usar el siguiente comando para unificar los archivos y crear uno solo en formato MD:

```
pandoc a-apertura.md b-desarrollo.md ../prototipos/analisis-de-la-calidad-de-los-microdatos-extraidos--7nfst.md.html c-cierre.md d-  
apendices.md -o tesis-pregrado.html --template=../plantillas/easy_template.html  
--toc Para adicionar o actualizar el PDF de La tesis, en la carpeta raíz del repositorio ejecutamos:
```

```
fossil.exe uv add capitulos/tesis-pregrado.pdf
```

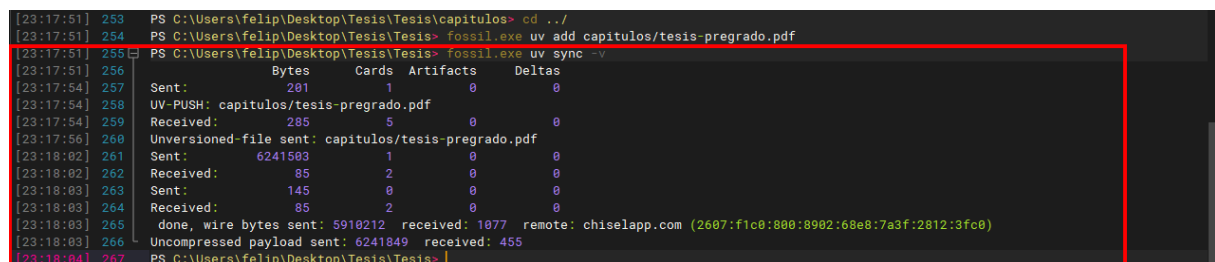


```
[23:15:51] 254 PS C:\Users\felip\Desktop\Tesis\Tesis\capitulos> cd ../  
[23:15:51] 255 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe uv add capitulos/tesis-pregrado.pdf  
[23:15:58] 256 PS C:\Users\felip\Desktop\Tesis\Tesis> |
```

**Figure 18:** Figura 1: Uso de los comandos

Para sincronizarlo con la versión remota hacemos:

```
fossil.exe uv sync -v
```



```
[23:17:51] 253 PS C:\Users\felip\Desktop\Tesis\Tesis\capitulos> cd ../  
[23:17:51] 254 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe uv add capitulos/tesis-pregrado.pdf  
[23:17:51] 255 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe uv sync -v  
[23:17:51] 256 Bytes Cards Artifacts Deltas  
[23:17:54] 257 Sent: 201 1 0 0  
[23:17:54] 258 UV-PUSH: capitulos/tesis-pregrado.pdf  
[23:17:54] 259 Received: 285 5 0 0  
[23:17:56] 260 Unversioned-file sent: capitulos/tesis-pregrado.pdf  
[23:18:02] 261 Sent: 6241503 1 0 0  
[23:18:02] 262 Received: 85 2 0 0  
[23:18:03] 263 Sent: 145 0 0 0  
[23:18:03] 264 Received: 85 2 0 0  
[23:18:03] 265 done, wire bytes sent: 5918212 received: 1077 remote: chiselapp.com (2607:f1c0:800:8902:60e0:7a3f:2012:3fc0)  
[23:18:03] 266 Uncompressed payload sent: 6241849 received: 455  
[23:18:04] 267 PS C:\Users\felip\Desktop\Tesis\Tesis> |
```

**Figure 19:** Figura 1: Uso de los comandos

Cuando este soncronizado se usa este comando para unificar los archivos en formato MD y pasarlo o convertirlo en PDF:

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

```
pandoc a-apertura.md b-desarrollo.md ../prototipos/analisis-de-la-
calidad-de-los-microdatos-extraidos--7nfst.md.html
c-cierre.md d-apendices.md -o tesis-pregrado.pdf --template eisvogel -
-pdf-engine=tectonic --toc -V lang=es-CO
```

Este comando va a unificar a todos los archivos nombrados y los convertirá en un archivo PDF, usando las herramientas Tectonic y Eisvogel Template.

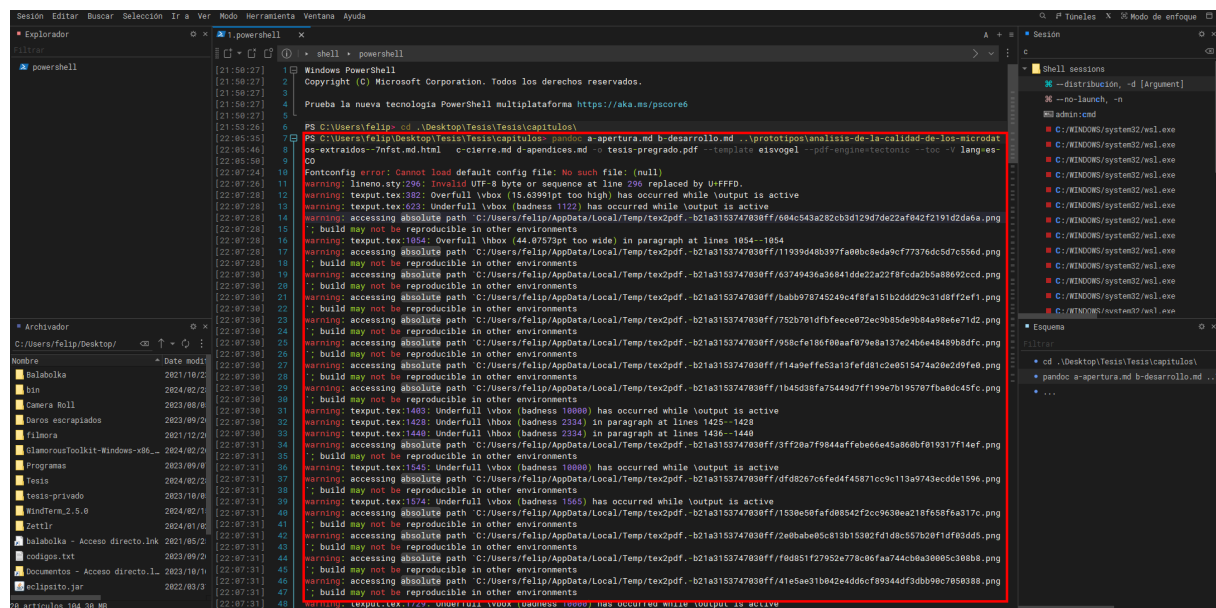


Figure 20: Figura 1: Uso de los comandos

Después se usa este comando para añadir el archivo al repositorio:

```
fossil.exe add .\trabajo-grado.md
```

Y por último se usa este comando para publicar el archivo añadido al repositorio:

```
fossil.exe commit -m "Nombre que desees"
```

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

```
[23:20:34] 252 PS C:\Users\felip\Desktop\Tesis\Tesis\capitulos> cd ../
[23:20:34] 253 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe uv add capitulos/tesis-pregrado.pdf
[23:20:34] 254 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe uv sync -v
[23:20:34] 255
[23:20:34] 256 Sent:          Bytes      Cards  Artifacts  Deltas
[23:20:34] 257 UV-PUSH: capitulos/tesis-pregrado.pdf
[23:20:34] 258 Received:     285        5        0          0
[23:20:34] 259 Unversioned-file sent: capitulos/tesis-pregrado.pdf
[23:20:34] 260 Sent:         6241503      1        0          0
[23:20:34] 261 Received:      85        2        0          0
[23:20:34] 262 Sent:          145        0        0          0
[23:20:34] 263 Received:      85        2        0          0
[23:20:34] 264 done, wire bytes sent: 5910212 received: 1077 remote: chiselapp.com (2607:f1c0:800:8902:60e8:7a3f:2812:3fc0)
[23:20:34] 265 Uncompressed payload sent: 6241849 received: 455
[23:20:34] 266 PS C:\Users\felip\Desktop\Tesis\Tesis> fossil.exe commit -m "actualizacion de los pantallazos"
[23:20:35] 267 Pull from https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
[23:20:36] 268 Round-trips: 1 Artifacts sent: 0 received: 0
[23:20:36] 269 Pull done, wire bytes sent: 455 received: 1739 remote: 2607:f1c0:800:8902:60e8:7a3f:2812:3fc0
[23:20:38] 270 New_Version: 3c63a0896d3cb8a0e1ea94c072cf16af90ddd565
[23:20:40] 271 Sync with https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
[23:20:41] 272 Round-trips: 1 Artifacts sent: 3 received: 0
[23:20:41] 273 Sync done, wire bytes sent: 4701 received: 319 remote: 2607:f1c0:800:8902:60e8:7a3f:2812:3fc0
[23:20:42] 274 PS C:\Users\felip\Desktop\Tesis\Tesis>
```

## **Adquisición de datos: Herramientas de Scraping**

En la investigación se utilizaron varias herramientas de scraping, que son programas o aplicaciones informáticos creados para recopilar datos de fuentes de Internet (en este caso, plataformas de redes sociales como Twitter). Recopilar información pertinente sobre las candidaturas de Gustavo Bolívar y Juan Daniel Oviedo a la alcaldía de Bogotá en Twitter/X es el objetivo principal de esta fase. Se elegirán cuidadosamente las herramientas de scraping adecuadas para los propósitos a fin de adquirir los datos. Estos programas extraerán datos de cuentas de Twitter, tweets, seguidores y cualquier otro dato pertinente relacionado con las elecciones.

Para garantizar que las conclusiones sean sólidas. Compararemos las herramientas. Es necesario analizar cómo funciona y sopesar sus ventajas e inconvenientes. Aprenderemos cómo funciona cada herramienta en términos de capacidad de recopilar y analizar datos, y la calidad de los datos extraídos de la plataforma de redes sociales. Nos enfocaremos en evaluar la precisión y fiabilidad de las herramientas, sin hacer referencia a la velocidad ni a la superación de limitaciones de la plataforma. Una vez obtenidos los datos pasaremos a su transformación. Para un análisis más sencillo, esto incluye limpiar y organizar los datos. Luego, escribimos historias con componentes tanto escritos como visuales. Estas narrativas permitirán representar eficazmente los hechos y ayudar a una mejor comprensión de los resultados.

La fase final de la investigación será el análisis del discurso. En esta sección, analizaremos los contenidos recopilados para encontrar patrones, tendencias y mensajes importantes relacionados con las campañas de Twitter/X de Gustavo Bolívar y Juan Daniel Oviedo. Este análisis nos ayudará a comprender mejor cómo se comparte la información en la plataforma y cómo afecta a la opinión pública y las decisiones de los ciudadanos. También nos brindará información útil sobre cómo se distribuye la información en el sitio.

### **Apify**

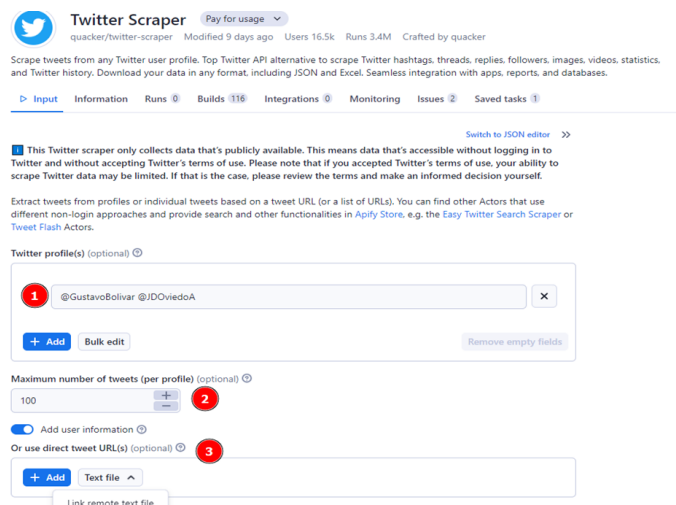
Apify es una plataforma flexible que proporciona herramientas para realizar operaciones de web scraping de forma eficaz y automática. "Twitter scraper" es uno de los muchos servicios que ofrece, que permite a los usuarios extraer automáticamente datos importantes de esta red social. Esto es particularmente útil para investigaciones, análisis de mercado y seguimiento de las tendencias de Twitter. Los usuarios del "Scraper para Twitter" de Apify pueden configurar parámetros particulares, incluidos términos de búsqueda, filtros de tiempo, propiedades de tweets e individuos específicos. El scraper se encarga de gestionar las llamadas a la API de Twitter una vez establecidos estos parámetros, recopilar los datos deseados y presentarlos de forma útil y estructurada. Esta estrategia técnica de Apify, junto con su "scraper para Twitter", ofrece una forma eficaz de automatizar la recogida de datos en Twitter, evitando la necesidad de realizar este procedimiento manualmente. Apify

agiliza el proceso de obtención de datos útiles de Twitter/X y los pone en manos de quienes tienen mayor conocimiento técnico al ofrecer una interfaz fácil de usar para personalizar y administrar la configuración de scraping.

A través de los enlaces mencionados en los tweets o los perfiles de las personas que estás viendo, Apify te permite revelar los detalles de los tweets. Pero existe una gran diferencia entre los modelos gratuitos y de pago. Por ejemplo, sería posible especificar fechas precisas o crear un motor de búsqueda personalizado de tweets utilizando un lenguaje de programación. Sin requerir el uso de una API, la plataforma ofrece datos sobre las características clave de los tweets, con:

- Cantidad de me gustas.
- ReTweets.
- Favoritos o guardados.
- Cantidad de comentarios.
- Cantidad de visitas.
- Enlace URL.
- Fecha y origen del Tweet.
- Descargar los datos.

La interfaz de Twitter Scraper luce de la siguiente manera:



**Figure 21:** Figura 1: Captura Apify

En la página principal se puede encontrar:

1. Configurar el perfil de un candidato le permite ingresar los nombres de los candidatos cuyos tweets desea recopilar o enlaces a sus perfiles. Al hacerlo, podrá recopilar datos pertinentes y concentrar los esfuerzos de recolección en cuentas particulares.

2. Se puede especificar la cantidad exacta de tweets que desea recopilar o se puede definir un máximo. Debido a su adaptabilidad, puede personalizar la colección para satisfacer sus necesidades, ya sea que esté buscando una descripción general o un grupo de datos en particular.
3. Tiene la opción de incluir enlaces a tweets particulares aquí. Esto es útil si desea recopilar datos precisos de tweets particulares, tuvo un impacto en los investigadores y desarrolladores (por ejemplo, extendió el tiempo de esta tesis más allá de lo habitual).

Además de eso, se pueden configurar:

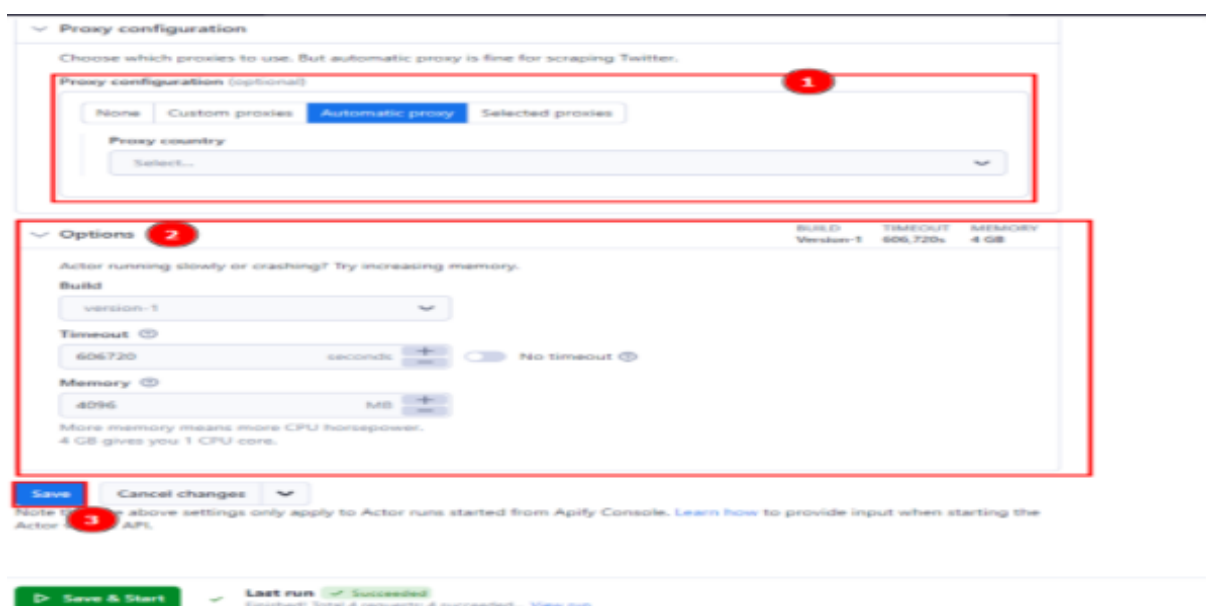


Figure 22: Figura 2: Captura Apify

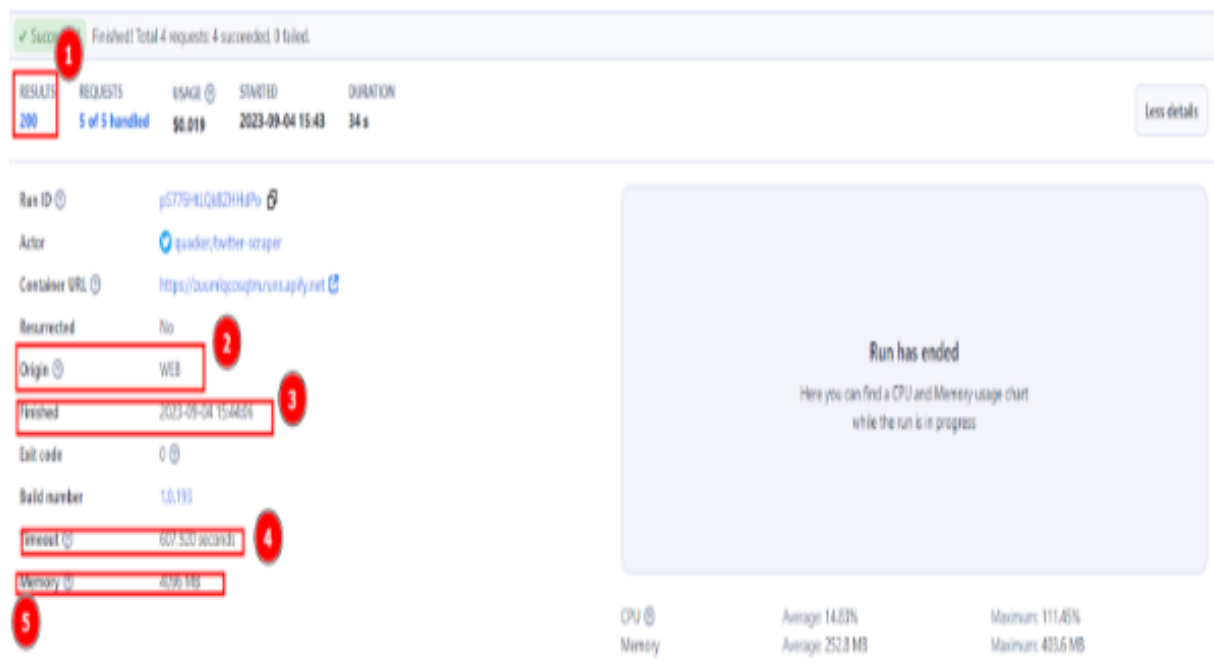
1. **Elección de las propiedades de Scraping:** Puede configurar propiedades específicas para su proceso de Scraping de Twitter en esta sección. Incluido: Ubicación del tweet: puede optar por recopilar solo tweets de una región determinada, lo cual es útil para monitorear eventos y tendencias locales. Tipos de tweets: puede optar por recopilar solo tipos específicos de tweets, como aquellos que son originales, retuiteados, mencionados o que contienen un hashtag o término en particular. Esto le permite concentrarse en extraer contenido particular. Incluya las URL exactas de los tweets que desea incluir en su colección seleccionando esta opción. Esto es útil para adquirir información precisa y contextual.
2. **Configuración para opciones avanzadas:** Puede cambiar los siguientes parámetros en la sección de opciones avanzadas: Se debe especificar la versión de la página de Twitter que desea utilizar para el scraping, ya que las páginas web pueden cambiar con el tiempo. Puede controlar durante cuánto tiempo la herramienta de raspado recopiló la información. Esto es útil para gestionar la duración del proceso y evitar retrasos innecesariamente largos. Límite en la recopi-



lación de datos: puede poner un límite a los datos que desea recopilar, lo cual es útil para evitar la descarga de grandes cantidades de datos. Asignación de memoria: debido a que puede afectar la rapidez con la que su sistema descarga datos, es crucial definir cuánta memoria desea dedicar a esta tarea.

3. **Comience a recolectar:** Simplemente elija "Iniciar" para iniciar el proceso de recopilación después de seleccionar todas las propiedades y opciones según sus especificaciones. El programa ahora comenzará a funcionar de acuerdo con sus comandos y recopilar datos de Twitter de acuerdo con la configuración que haya definido.

Se proporcionarán detalles importantes sobre el procedimiento de recopilación de datos de búsqueda y tweets en la nueva parte que será visible en la parte inferior de la página de recopilación de datos del usuario de Prefin, incluida la siguiente información:



**Figure 23:** Figura 3: Captura Apify

1. La cantidad total de tweets recopilados exitosamente durante el procedimiento se muestra a continuación en Resultados o Número de tweets recopilados. Esto le ayudará a comprender cuántos datos ha recopilado.
2. Fuente de la encuesta: Al indicar la fuente de la encuesta en este campo, será posible rastrear el origen o el objetivo de la recopilación de datos.
3. Fecha y hora en que se iniciará el proceso de recogida de datos: Se le notificará cuando lo haga. Esto es importante para el seguimiento temporal y de registro de la operación.
4. La cantidad de tiempo necesaria para finalizar la recopilación de datos se mostrará según el

tiempo utilizado para recopilar la información. Esto es crucial para determinar la efectividad y duración de la tarea.

5. Memoria utilizada para la recopilación: con esta información, se puede determinar cuánta memoria se necesita para completar la recopilación. Esto es crucial para la gestión de recursos y para garantizar que haya suficiente RAM disponible para las próximas tareas.

Al final de la sección se proporciona una recopilación de tweets sobre los dos candidatos, clasificados del más relevante al menos importante. Para cada tweet de esta lista, se mostrará la siguiente información detallada:

1	2	3	4	5	6	7	8
tweet_text	Picture media_url	Replies reply_count	Retweets retweet_count	Favourites favorite_count	Views view_count	URL	Date created_at
Medida de aseguramiento al expresid		3992	11500	47958	undefined	https://twitter.com/GustavoBo...	2020-08-04 13:17:35
Brutal la lavata de imagen q los ma		1848	12112	35244	undefined	https://twitter.com/GustavoBo...	2020-08-05 13:16:06
ATENCIÓN POR EL TRACING: 39 VO		871	7500	34726	undefined	https://twitter.com/GustavoBo...	2020-09-02 23:32:27
#DelCongresoEspere un mes de vaca		2362	7367	34912	undefined	https://twitter.com/GustavoBo...	2020-06-24 12:59:32
Jesús Uribe tiene todo el derecho d		1990	7800	33092	undefined	https://twitter.com/GustavoBo...	2020-04-21 19:00:41
Maria Fernanda Cabal cree que la ma		1989	10070	32447	undefined	https://twitter.com/GustavoBo...	2020-06-25 18:02:25
¿La cadena perpetua se va a estren		2651	8909	32059	undefined	https://twitter.com/GustavoBo...	2020-06-24 11:26:17
Postulo mi nombre como Presidente d		4243	8572	35326	undefined	https://twitter.com/GustavoBo...	2022-06-21 13:51:45
No sido elegido Presidente de la Co		3769	5619	33287	undefined	https://twitter.com/GustavoBo...	2022-07-26 17:22:42
Epa Colombia por romper un tablero		1520	12767	31487	undefined	https://twitter.com/GustavoBo...	2021-08-12 13:17:22

Figure 24: Figura 4: Captura Apify

1. El texto completo del tweet, que puede incluir enlaces, comentarios, opiniones u otra información publicada por el usuario.
2. Imagen (si está disponible): Si el tweet tiene una imagen asociada, se mostrará para que puedas ver visualmente el contenido multimedia compartido.
3. Una medida de la actividad y el discurso que rodea a un tweet es la cantidad de comentarios o respuestas que ha recibido de otros usuarios.
4. Retweets: Cuántas veces un tweet fue compartido por otros usuarios, indicando qué tan ampliamente fue visto en la plataforma.
5. Número de Me gusta: la cantidad de veces que los usuarios hicieron clic en el botón "Me gusta" junto a un tweet para mostrar su apoyo y popularidad.
6. Si el sitio web ofrece información sobre el número de visitantes o vistas de tweets, esta información podría estar presente. Esto podría resultar útil para evaluar la visibilidad del contenido.
7. URL: el enlace al tweet en particular, que le permite acceder a él directamente en Twitter y obtener más información sobre él o interactuar con él.
8. La hora y fecha precisas en que se publicó el tuit permiten contextualizar la información a la luz

de su naturaleza temporal.

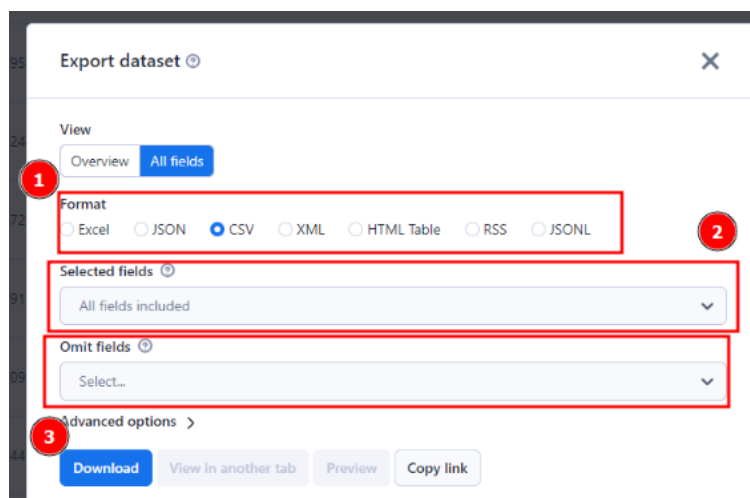
Se puede encontrar la siguiente información en la sección "All Fields", que proporciona información adicional y completa sobre los tweets recopilados:

Conversation ID	Created_at	Favorite_count	Full_text	Hashtag ID	Is_gilded	Is_quote_tweet	Is_retweet	Is_truncated	Media	Quote_count	Quote_tweet	Reply_count	Retweet_count
129071467029357319	2023-05-12T17:20:04	47950	Medico de aseguramiento al deposit... (truncated)	2 items		X	X	X	0 items	740	undefined	3392	11563
1291811522024812429	2023-05-13T16:06:00	22284	Realiza la lectura de imagen e... (truncated)	0 items		X	X	X	0 items	401	undefined	1340	12112
1391311460672515145	2023-05-10T23:27:00	38726	ATENCION: PCR LL (BAC/NEL: 59 VU) <> (truncated)	1 item		X	X	X	0 items	427	undefined	871	7163
1540394204892366400	2023-05-12T09:32:24	84910	#RealIngredientes un mas de vaca... (truncated)	1 item		X	X	X	0 items	386	undefined	2567	7547
1212749122819547022	2023-05-11T23:41:21	22382	Just Unlike Bone Inks el... (truncated)	1 item		X	X	X	0 items	422	undefined	1392	7022
1376168874996616108	2023-05-10T20:25:15	51841	Maria Hernandez (2023... (truncated)	0 items		X	X	X	2 items	946	undefined	1989	10816
1375817897870168035	2023-05-11T20:47:24	52069	¿La cadena productiva en... (truncated)	0 items		X	X	X	0 items	710	undefined	3051	8880
1529310108114571264	2023-05-13T14:43:00	33329	Postula su nombre como... (truncated)	0 items		X	X	X	0 items	1005	undefined	4242	8372

Figure 25: Figura 5: Captura Apify

1. **ID:** Código o número de identificación especial del tweet. Cada tweet tiene un número de identificación único que puede usarse para rastrear con precisión las entradas.
2. **hashtag count:** El número total de hashtags utilizados en el tweet. En las redes sociales, los hashtags (palabras clave seguidas del signo "#") se utilizan para agrupar y etiquetar material.
3. **Exportación de datos:** Simplemente haga clic en el botón "Exportar" en la parte inferior de la página para obtener los datos recopilados. Con el uso de esta herramienta, puede guardar datos en un archivo o formato particular, como un archivo Excel o CSV, para un estudio o referencia posterior.

Tiene una variedad de opciones adaptables en la página para personalizar la exportación de datos, de modo que puede ajustar la descarga del tweet según sus necesidades:



**Figure 26:** Figura 6: Captura Apify

1. Seleccione la opción Formato de exportación para especificar el tipo de archivo para la descarga de datos. Las opciones más utilizadas son Excel, JSON, CSV, XML, tablas HTML, RSS y JSONL. Podrás seleccionar el formato que mejor se adapte a tu proceso de análisis o visualización de datos porque cada uno tiene sus propios beneficios y aplicaciones.
2. **Campos seleccionados (Selected Fields):** Puedes definir uno o más requisitos que deben cumplir los tweets para poder ser incluidos en la exportación en esta sección. Esto hace posible filtrar los tweets que no se ajustan a criterios particulares y descargar solo aquellos que sí lo hacen, lo cual es útil para técnicas de análisis más precisas.
3. **Omitir campos:** puede establecer criterios particulares que los tweets no deben cumplir para ser omitidos de la exportación en la sección "Omitir campos". Esto hace posible eliminar los tweets que no cumplen con requisitos particulares, lo cual es útil para deshacerse de datos innecesarios o innecesarios del conjunto de datos exportados.

## Twitter Scraper

El término "Twitter Scraper" se refiere a un programa que extrae mecánicamente archivos y tweets de Twitter. Un "Twitter Scraper" es una extensión creada específicamente para recopilar información de perfil de Twitter y tweets directamente desde el navegador Chrome, en el contexto de una extensión de Google Chrome.

**Interacción con Twitter:** La extensión se puede activar automáticamente cuando visita una página de perfil de Twitter o accede a Twitter. Una extensión simula acciones que normalmente realizamos manualmente, como hacer clic en enlaces o desplazarse a la página, mediante el uso de tecnologías automáticas.

**Extracción de datos:** Cuando la extensión está en una página de perfil o en los resultados de búsqueda, puede buscar y extraer información pertinente como nombres de usuario, nombres reales, biografías, imágenes de perfil, números de suscriptores, tweets, etc. Estos datos se recopilan y guardan localmente o se procesan según las preferencias del usuario.

**Almacenamiento y presentación:** Los datos extraídos se pueden guardar de forma legible en la extensión o en la PC del usuario. Algunas extensiones también pueden proporcionar los datos recopilados de forma ordenada y comprensible en la interfaz del navegador.

**Cambios en la estructura del sitio:** Las extensiones de scraping están determinadas por la estructura y el diseño de la página web. Si el diseño o la estructura de Twitter cambian, la extensión puede dejar de funcionar correctamente hasta que se actualiza para reflejar los cambios.

Para que la extensión funcione correctamente en Google Chrome, primero debe darle permiso para conectarse. Los siguientes pasos necesarios:



Figure 27: Figura 7: Captura Twitter Scraper

1. **Escriba su nombre de usuario o perfil aquí:** En la barra de búsqueda de la extensión, escriba el nombre de usuario o ID del perfil que desea buscar. Verifique que este inicio de sesión esté activo y sea legítimo en la plataforma que está utilizando.
2. **Indique una duración:** Díganos cuánto tiempo planea dedicar a recopilar información. Su recopilación de datos será más exhaustiva porque abarca un período de tiempo mayor si invierte más tiempo en ella. Configure esta opción según sea necesario.
3. **Contiene el año y mes de inicio:** Establezca el año y el mes desde el que la extensión debe comenzar a recopilar datos. La recogida se realizará entre la fecha de inicio que elijas y el día

actual. Los valores que especifique para el período de tiempo que desea estudiar deben ser precisos.

4. **Seleccione "Scrape Now"**: Para iniciar el proceso de recopilación de datos, seleccione la opción "Scrape Now" después de personalizar todos los parámetros según sus preferencias. A partir de este momento, la extensión comenzará a recopilar datos del perfil del usuario de acuerdo con las pautas.

La extensión mostrará al usuario una nueva página con amplia información sobre los tweets de ese perfil una vez finalizada la recopilación de datos. Puede encontrar los siguientes detalles en esta página:

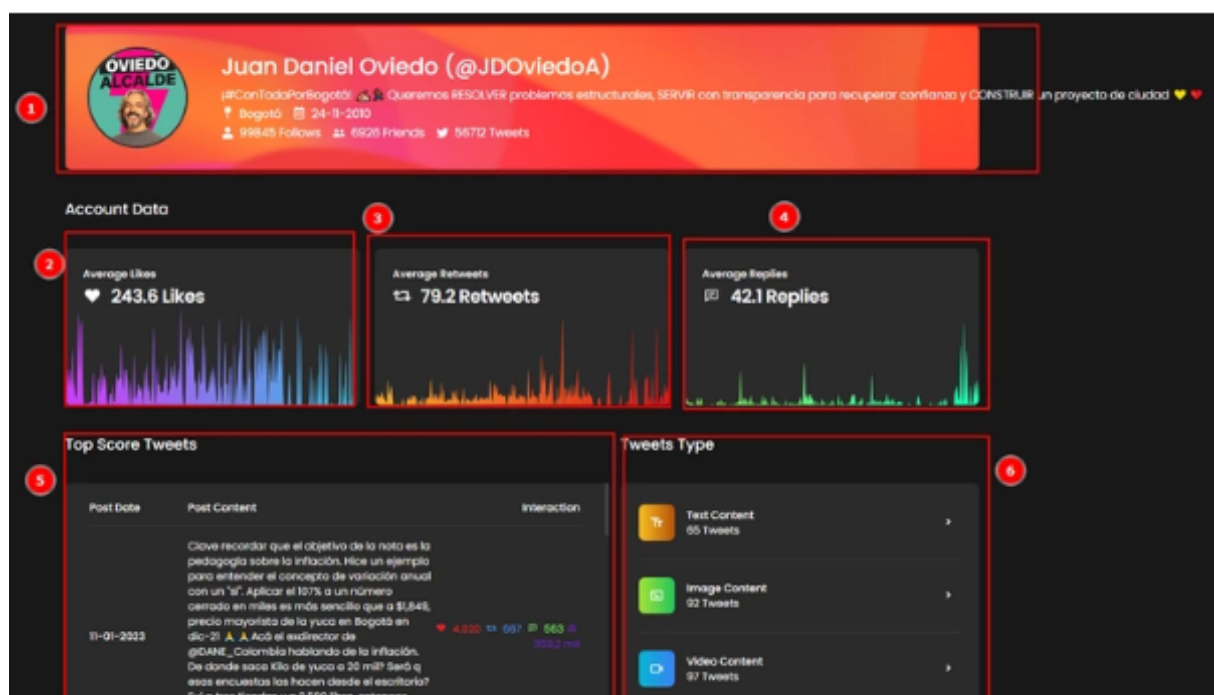


Figure 28: Figura 8: Twitter Scraper

1. **Información básica del perfil de usuario:** esta parte brindará la información fundamental del perfil de un usuario, como el nombre de usuario, la descripción de la cuenta, la fecha de creación y otra información pertinente.
2. **Número de Me gusta:** el perfil del usuario indicará el número total de Me gusta que ha acumulado en todos sus tweets. Esto ilustra lo apreciados y aceptados que son tus productos.
3. **Número de "Retweets":** Esto mostrará cuántas veces los tweets del perfil han sido retuiteados por otros usuarios, mostrando cuán leídas y relevantes son tus publicaciones.
4. **Número de comentarios o respuestas:** muestra el número total de comentarios o respuestas producidas por los tweets del perfil, mostrando la discusión y actividad en torno a sus publica-

ciones.

5. **Lista de tweets con más interacciones:** esta sección muestra una lista de cada tweet que ha recibido más me gusta, tweets y comentarios. Luego podrá ver qué tweets son los más importantes y populares en el perfil.
6. **Lista de tipos de tweets:** los tweets se agrupan según el tipo de contenido que contienen, incluidos texto, fotos y videos. Puedes ver cómo la persona usa varios tipos de contenido en su perfil al hacer esto.

Encontrará herramientas destinadas a facilitar la búsqueda y clasificación de tweets en la sección “Listas de publicaciones”. Estos recursos incluyen:

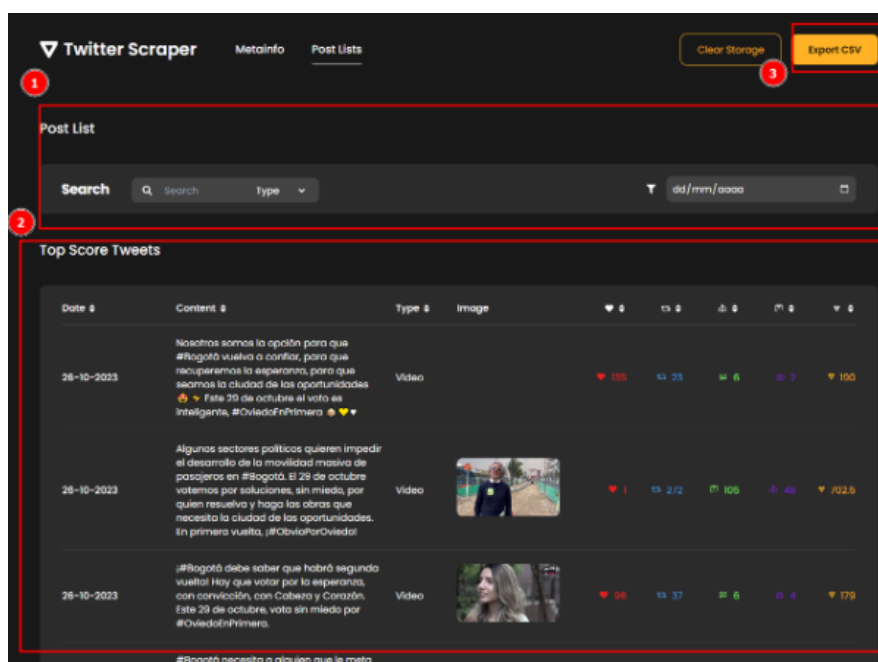


Figure 29: Figura 9: Twitter Scraper

1. **Búsqueda avanzada:** Esta función facilita la búsqueda de contenido. Puedes buscar por tipo de contenido o usar palabras clave específicas relacionadas con el tweet que estás buscando.
2. **Detalles de la publicación:** Puede encontrar detalles detallados sobre cada tweet aquí. La fecha de publicación, el tipo de contenido del tweet, el formato del tweet (por ejemplo, texto, imagen, enlace, etc.), la cantidad de me gusta, retweets, reacciones y comentarios recibidos se encuentran entre esta información
3. **Descarga de datos:** También puede usar el proceso de “scraping” para descargar los datos obtenidos en formato CSV. Esto le permite realizar un seguimiento y análisis posteriores de los tweets.

## Squawker

FOSS, que en inglés significa “Free and Open Source Software”, es una plataforma que destaca por ser una biblioteca de software instalable. Antes de poder acceder a la aplicación “Squawker”, primero debes descargar otra aplicación llamada “F-Droid”. software; gratis. Su principal objetivo es brindar a los usuarios de Android acceso a un repositorio de aplicaciones móviles de código abierto y gratuito para que puedan navegar, descargar y usar. A diferencia de las tiendas de aplicaciones convencionales, F-Droid promueve software que se adhiere a los principios del software libre, lo que permite a los usuarios utilizar, estudiar, cambiar y redistribuir el código fuente de los programas.

El cerebro de F-Droid está contenido en el cliente, una aplicación de Android que instalas y utilizas para acceder a tu colección como un portal. El cliente F-Droid permite a los usuarios buscar, explorar y seleccionar entre una variedad de programas FOSS. Además, instalar estas aplicaciones directamente en el teléfono inteligente es sencillo y no requiere visitar tiendas especializadas como Google Play Store.

La capacidad del cliente F-Droid para buscar actualizaciones del software instalado desde su repositorio y notificarlas a los usuarios es una característica crucial. Gracias a esto, los usuarios siempre tendrán acceso a la información y la configuración de seguridad más recientes.

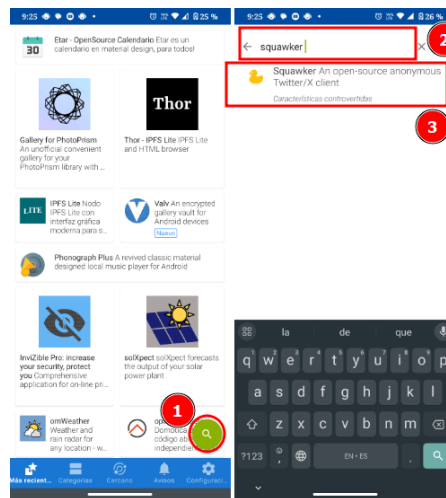


**Figure 30:** Figura 10: Squawker

En la esquina inferior derecha de la pantalla se encontrará un ícono de lupa, simplemente selecciónelo y escriba el nombre del software deseado, en este ejemplo Squawker. El nombre del desarrollador y una imagen o logotipo representativo que lo acompaña aparecerán en la pantalla cuando el programa haya realizado una búsqueda automática en la base de datos de F-Droid. Esto hace que encontrar e identificar el software que desea instalar es bastante sencillo.



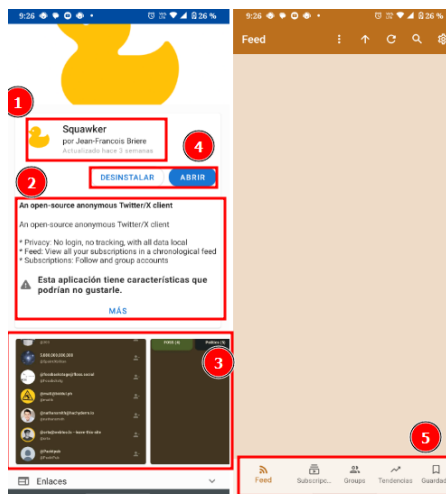
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 31:** Figura 11: Squawker

1. Opción de búsqueda.
2. Buscador.
3. Aplicativo (Squawker).

Al momento de seleccionar la aplicación va a aparecer una nueva interfaz con el contenido general y explícito del aplicativo:



**Figure 32:** Figura 12: Squawker

1. Nombre del aplicativo y creador.
2. Requerimientos o condiciones que se necesita para instalar la aplicación.
3. Imágenes para mostrar el diseño de la aplicación.
4. Opción de descarga.

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

Después de descargar la aplicación y ejecutarla, va aparecer la interfaz o apartado principal de la aplicación donde se encontrar todas las opciones disponibles para iniciar la extracción de información o datos de Twitter/X como:

**Feed:** Al seleccionar esta opción, se le dirigirá al área donde podrá navegar y seguir las actualizaciones de Twitter o la fuente de información necesaria.

**Suscripciones:** Para recibir notificaciones o actualizaciones pertinentes, puede administrar sus suscripciones a personas, temas o cuentas particulares aquí.

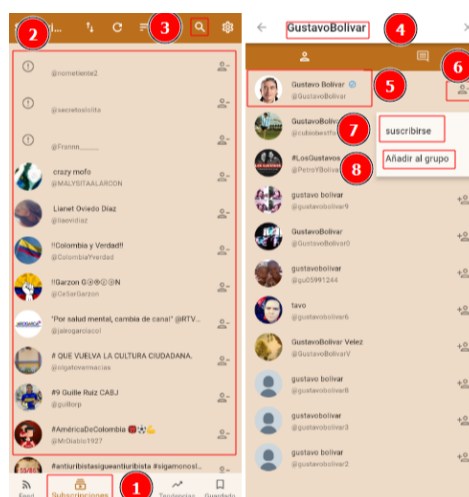
**Grupos:** Esta opción te permitirá crear varias categorías o información relacionada si el programa permite agrupar elementos en categorías.

**Tendencias:** Para mantenerse al día con los temas candentes, consulte las tendencias más recientes en Twitter u otra fuente de datos relevante.

**Guardado:** Si desea conservar o marcar elementos específicos para su uso posterior, puede utilizar esta opción para guardar o marcar elementos relevantes.

Puede encontrar una lista o directorio de perfiles de Twitter/X en la sección "Suscripción". Simplemente haga clic en el ícono de la lupa en la parte superior de la pantalla para encontrar un perfil determinado. Se debe ingresar el nombre de usuario del perfil que desea ver. En tu directorio, el programa localiza automáticamente un perfil con tu foto de perfil, y si es famoso mostrará el signo de verificación correspondiente.

Simplemente haga clic en el símbolo de persona en el lado derecho del perfil para acceder a él. Aparecerán dos opciones "suscribirse" y "agregar al grupo" estarán disponibles. Cuando te suscribes, los tweets de los perfiles que siguen aparecerán en tu feed y podrás navegar por el perfil del usuario.

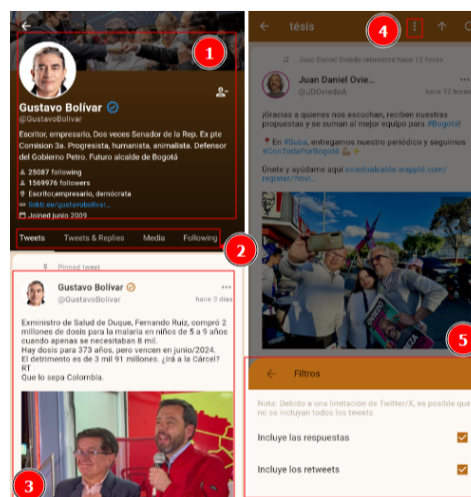


**Figure 33:** Figura 13: Squawker

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

1. Apartado de Suscripciones.
2. Catálogo de perfiles.
3. Opción de búsqueda.
4. Nombre del perfil de usuario.
5. Perfil del usuario solicitado.
6. Apartado de suscripción o añadir a grupo del perfil.
7. Opción de suscribirse.
8. Opción para añadir grupo.

Seleccione el perfil que desea revisar para ver información general y específica del usuario. El software automáticamente lo dirige a una nueva interfaz donde puede ver todos los detalles del perfil del usuario. Encontrará varias secciones en este enlace para ayudarlo a descubrir contenido. La cantidad de contenido e información disponible aumenta aún más gracias a las opciones de filtro, que también incluyen respuestas y retuits de tweets individuales.



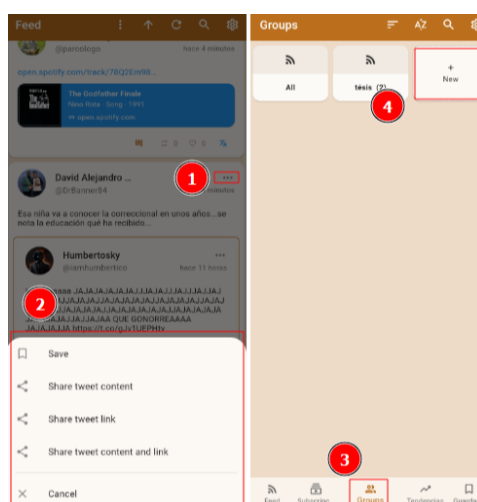
**Figure 34:** Figura 14: Squawker

1. Información de perfil de usuario (general y específica).
2. Apartados para facilitar la búsqueda de un contenido en específico. con una característica en específica.
3. Apartado para habilitar la opción de filtros.
4. Opción para activar los filtros.
5. Filtros para limitar los Tweets.

Para guardar o compartir el Tweet se selecciona los tres puntos horizontales que se pueden encontrar en la esquina superior derecha del tweet. Al hacer clic en estos tres puntos se abrirá un menú con opciones adicionales relacionadas con ese tweet. Opciones como "Guardar", "Compartir contenido

de Tweet”, ”Compartir Enlaces de Tweet” y ”Compartir contenido y enlaces de Tweet” aparecerán cuando presiones los tres puntos. Para guardar el tweet, utilice la opción ”Guardar”.

Simplemente puedes acceder a un tweet que hayas guardado más tarde. Puedes organizar y categorizar tus tweets guardados en el área de grupos. En esta sección, puedes iniciar tu propio grupo. Puedes crear grupos para tus tweets favoritos, tweets más educativos o cualquier otra categoría que te guste. Simplemente seleccione la opción de ”Agregar al grupo” dicha opción que ya mencionamos con anterioridad. Desde allí se puede elegir el grupo en el que desea conservar esos tweets. Esto le facilita modificar sus tweets para adaptarlos a sus intereses y requisitos.



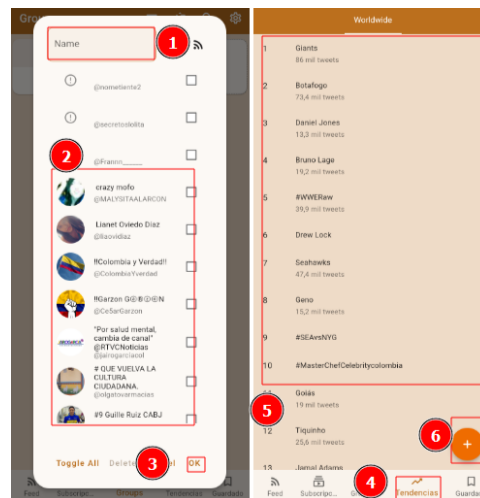
**Figure 35:** Figura 15: Squawker

1. Apartado para habilitar las opciones de los Tweets.
2. Opciones de guardar y compartir del Tweet.
3. Apartado de grupos.
4. Opción para crear grupos,

Se mostrará una sección que le permitirá establecer un nuevo grupo una vez que elija ”New”. Puedes ingresar el nombre que deseas darle al grupo en el campo ”Name” en la parte superior. Los perfiles que desea incluir en el grupo se mostrarán en forma de listado en la parte inferior. Una vez que haya terminado, haga clic en ”OK” para completar el proceso y crear el grupo.

En el apartado de ”Tendencias” se encontrará una lista de hashtags, palabras o temas que actualmente son tendencia en la red social y el símbolo de ”+” que aparece en la parte inferior a mano derecha es un filtro para las tendencias.

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



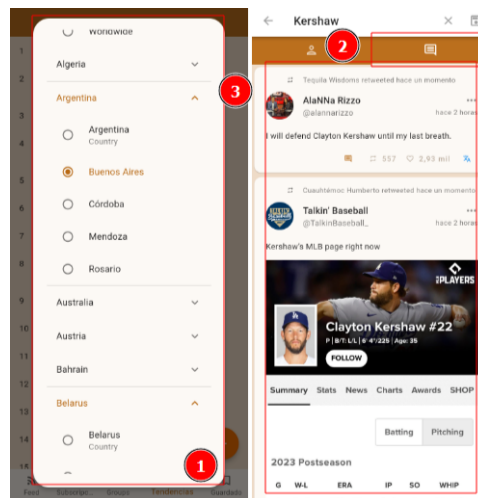
**Figure 36:** Figura 16: Squawker

1. Apartado para nombrar el grupo.
2. Listado de perfiles de usuarios.
3. Opción de crear el grupo.
4. Apartado de Tendencias.
5. Listado de tendencias.
6. Opción de Filtros.

Al elegir la opción de filtro, se puede acceder a una sección con filtros basados en la ubicación, lo que simplifica la identificación de patrones particulares en su región. El software lanzará inmediatamente una nueva parte con una lista de ítems adicionales una vez que elijas el lugar y la tendencia que te interesa.

Podrás ubicar todos los medios o publicaciones sobre la tendencia que seleccionaste en la primera sección. En esta sección se puede encontrar una selección de artículos, imágenes y vídeos relacionados con la tendencia en cuestión. Luego puede navegar y profundizar en las discusiones y el contenido creado por otros usuarios que estén interesados en esa tendencia en particular.

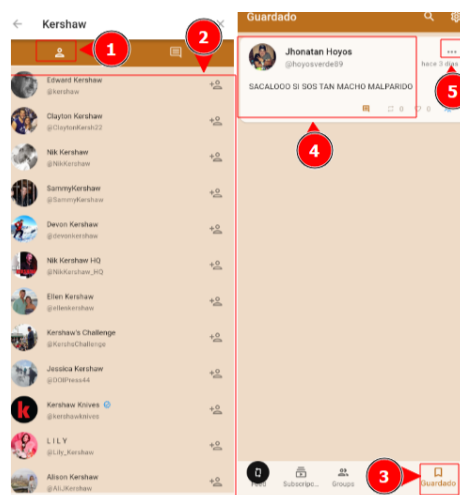
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 37:** Figura 17: Squawker

1. Filtro por ubicación.
2. Apartado de tendencia de contenido.
3. Contenidos o publicaciones relacionado con las tendencias.

En la segunda sesión se encontrará un listado de perfiles de usuarios con la opción de seguir dicho perfil, que estén relacionados con la tendencia seleccionada. En el apartado de “Guardado” se encontrarán todos tweets guardados por el usuario, en la opción Que tiene 3 puntos, ubicado en la parte superior a mano derecha del tweet, habilitar un apartado con varias opciones, “Unsave”, compartir el contenido, compartir enlace y la opción de compartir el contenido del tweet y su enlace.



**Figure 38:** Figura 18: Squawker

1. Sección dos.

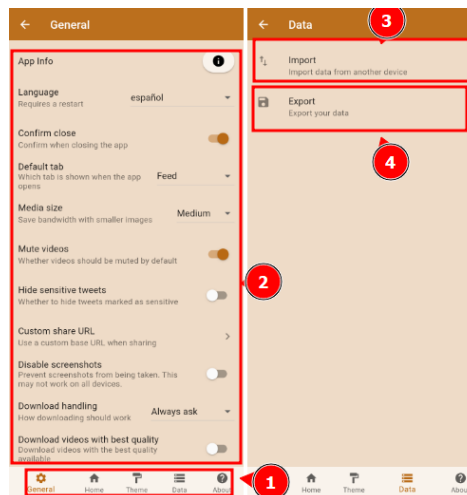
2. Listado de usuarios de perfiles.
3. Apartado de guardado.
4. Tweets guardados.
5. Apartado de opciones.
6. Opcion de configuración.

El usuario es llevado a una nueva interfaz después de elegir la opción de configuración en la aplicación, donde puede elegir entre una variedad de opciones para modificar varias características, incluido el diseño, las funcionalidades y la capacidad de importar y exportar datos. Se ofrecen cinco nuevas opciones para que las examine y modifique en la parte inferior de esta pantalla. "General" es la primera opción y comprende configuraciones relacionadas con la configuración del sistema de la aplicación. Los usuarios pueden personalizar su experiencia de usuario estableciendo preferencias y configuraciones generales aquí.

El usuario puede elegir qué secciones o preguntas quiere ver al iniciar sesión en la aplicación utilizando la sección "Home", que es la siguiente sección. Gracias a esto, tienen más control sobre el material de la página de inicio, lo que también simplifica la personalización de su experiencia.

Debido a que permite a los usuarios alterar la estética visual de la aplicación, la parte "Theme" es una característica con una apariencia visual agradable. Esto implica que los clientes pueden modificar la apariencia estética de la interfaz de la aplicación de acuerdo con sus gustos particulares. La parte "About" de la aplicación contiene detalles cruciales, incluida la versión actual y las licencias relacionadas, así como una ubicación para que usted informe cualquier problema o error que pueda encontrar. Los usuarios que quieran obtener más información sobre el software o ponerse en contacto con el personal de soporte encontrarán útil esta función.

La sección "Datos" ofrece dos alternativas sencillas al final. Los usuarios pueden importar datos desde la memoria de su dispositivo usando la primera opción, lo que simplifica la obtención de información crucial en la aplicación. Los usuarios pueden descargar tweets específicos u otra información utilizando la función de exportación de datos de la segunda opción, que luego pueden guardar o distribuir de varias maneras.



**Figure 39:** Figura 19: Squawker

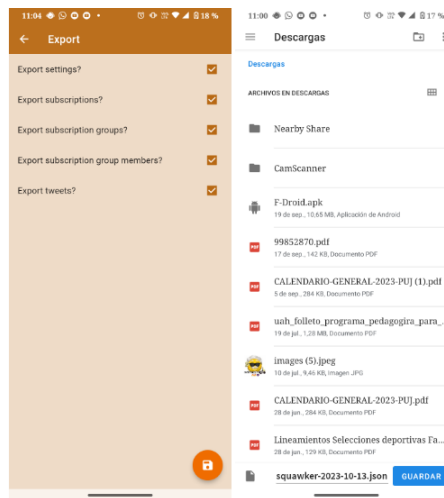
1. Secciones de configuración.
2. Opciones de configuración.
3. Opción de importar datos.
4. Opción de exportar datos.

Serás dirigido a una sección con numerosas opciones y recursos para descargar tus datos cuando elijas la opción "Exportar" en el programa. Hay varias opciones para exportar el tipo de dato o datos sobre el tweet seleccionado como "Export settings", "Export subscriptions", "Export subscription group", "Export subscription groups members" y "Export tweets". Cuando los datos estén preparados para la descarga, aparecerá un icono con un símbolo de disquete en la parte inferior después de haber especificado los componentes o funciones que desea incluir en la descarga de datos.

Al hacer clic en este botón, el programa abrirá una nueva pestaña donde podrá elegir la ubicación de almacenamiento de su dispositivo. Ahora puede elegir la ubicación donde desea guardar el archivo. Se utilizará el formato JSON para almacenar el archivo creado.



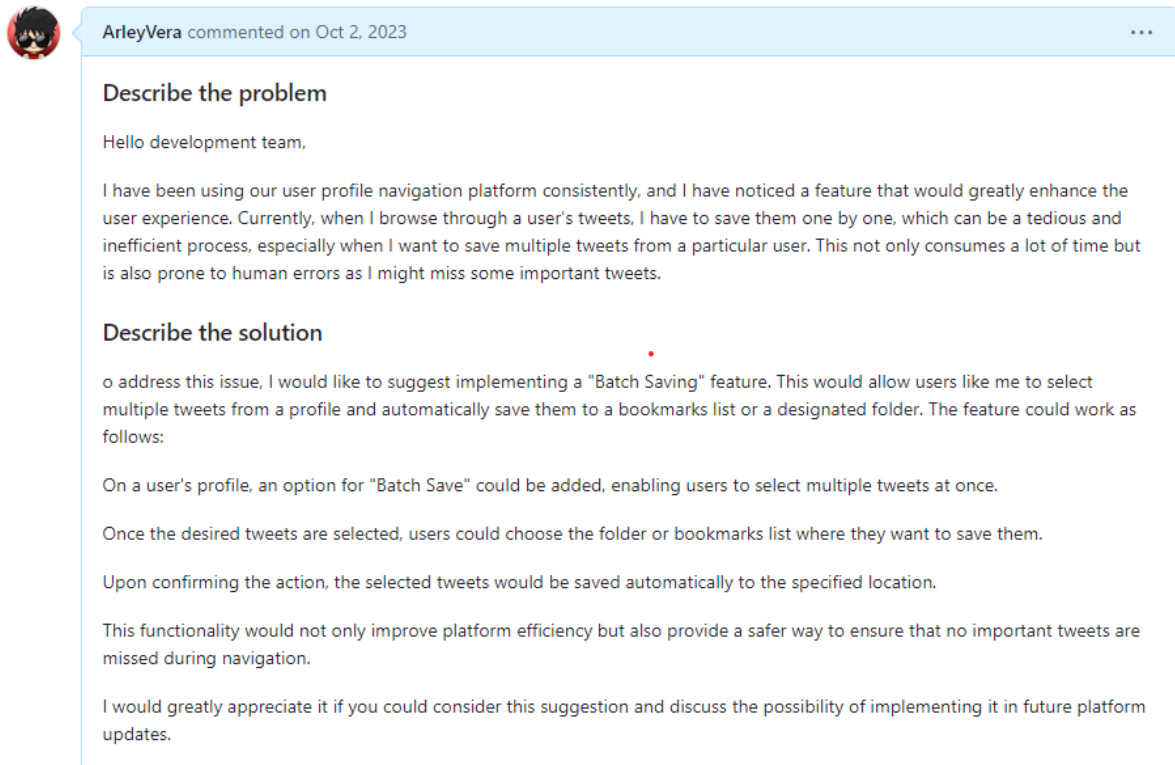
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 40:** Figura 20: Squawker

Se descubrió una debilidad en la aplicación Squawker durante el estudio para el trabajo de guardo. La falta de una función “Batch Save” para explorar perfiles. Se tomó la decisión de abrir una línea de comunicación con los desarrolladores vía GitHub por medio de los comentarios de dicho aplicativo.

Se describieron los problemas que enfrentan los usuarios cuando tienen que guardar trinos uno a uno en un comentario detallado, enfatizando la ineficiencia y la propensión a errores que esto causa. Se sugirió una solución práctica, la cual consistió en agregar una función “Batch Save” que permite a los usuarios elegir y almacenar varios trinos en una carpeta o lista de favoritos a la vez.



**Figure 41:** Figura 21: Squawker

**Tabla de comparación:**

<b>Aspecto</b>	<b>Apify</b>	<b>Twitter Scraper (Extensión de Chrome)</b>	<b>Squawker</b>
Tipo de Plataforma	Plataforma de web scraping automatizado	Extensión de navegador	App para Cell Phone (Android)
Funciones Principales	Extracción de datos de Twitter de manera automatizada	Extracción de datos de Twitter desde el navegador	Extracción de datos de Twitter desde la APP
Estrategia de Scraping	Búsqueda: <ul style="list-style-type: none"> <li>• Perfil</li> <li>• Hashtag</li> <li>• Palabra clave</li> <li>• Time para generar una consulta</li> <li>• Url</li> </ul>	Búsqueda: <ul style="list-style-type: none"> <li>• Perfil</li> <li>• Periodo de tiempo</li> <li>• Time para generar una consulta</li> <li>• Atributo para delimitar fechas o un periodo de tiempo</li> </ul>	Búsqueda: <ul style="list-style-type: none"> <li>• Perfil</li> <li>• Territorio</li> </ul>
Almacenamiento de Datos	Se almacenan los datos de la web de la página como un historial	No los almacena	Se almacena en el apartado de Guardar
Formato de exportación de Datos	(Excel, JSON, CSV, XML, HTML Table, RSS, JSONL)	CSV	J.son

<b>Aspecto</b>	<b>Apify</b>	<b>Twitter Scraper (Extensión de Chrome)</b>	<b>Squawker</b>
Conocimientos Técnicos	No requiere de algún conocimiento técnico, lo único que pide es un correo electrónico (Gmail) para acceder a la página.	Requiere autorización y configuración en el navegador	No requiere de algún conocimiento técnico
Información Disponible	Cantidad de me gusta, Retweets, Favoritos, Comentarios, Visitas, Enlace URL, Fecha y origen del Tweet, Descarga de datos	Información de perfil de usuario, Cantidad de "Me gusta," "Retweets," Comentarios, visitas y Listado de tweets con más interacciones	Información de perfil de usuario, Cantidad de "Me gusta," "Retweets," Comentarios, visitas y Listado de tweets con más interacción y tendencias
Consideraciones			La aplicación deja de funcionar constantemente La opción de descargar es difícil de encontrar al inicio No es muy claro el uso de sus herramientas

<b>Aspecto</b>	<b>Apify</b>	<b>Twitter Scraper (Extensión de Chrome)</b>	<b>Squawker</b>
Interfaz de Usuario	La interfaz puede ser pesada para un usuario que no tenga conocimiento alguno sobre técnicas de Scrap o sobre el sitio web y álabes puede ser poco digerida visualmente para los usuarios debido a tantas opciones o herramientas que muestra la página web	La interfaz es intuitiva para el usuario debido a que no tiene muchos apartados y los que tiene son muy claros para usarlos, también la interfaz es muy llamativa y es muy clara para su usabilidad para él usuario.	La interfaz es intuitiva pero muy básica, no tiene un diseño llamativo y sus herramientas como de importación, exportación y ajustes de sistemas no son fáciles de encontrar o no son muy claros para usar
Acceso a Datos Adicionales	Proporciona información adicional como ID y cantidad de Hashtags	Proporciona información básica del perfil y tipos de tweets	Proporciona información básica del perfil

## Análisis de la calidad de los microdatos extraídos

La siguiente narrativa de datos examina la información obtenida a través de la extracción de datos.

El objetivo es evaluar el nivel de detalle proporcionado por diversas técnicas de “scrapping”, específicamente en relación con los tweets, utilizando los metadatos incluidos en cada conjunto de datos extraído para definir los dos dimensiones del análisis que utilizamos para la investigación:

Dimensionalidad: es decir la cantidad de datos, referidos a distintos elementos (autor, hashtags, media, etc) que proporciona la fuente de datos.

2.Densidad: es decir la cantidad de datos que se encuentran agrupados por cada dimensión (cuánta información hay sobre el autor, cuánta sobre los hastags, cuanata sobre los media, etc).

### Prerrequisitos

Instalaremos Data Frame para Pharo

```
EpMonitor disableDuring: [
  Metacello new
    baseline: 'DataFrame';
    repository: 'github://PolyMathOrg/DataFrame/src';
    load ].
```

El proyecto DataFrame de Pharo proporciona herramientas para la gestión y el análisis de datos estructurados. Admite operaciones de acceso, modificación, enumeración, aritméticas y estadísticas en colecciones de datos. El paquete Pharo DataFrame ofrece métodos para detectar, eliminar y reemplazar valores nulos en una colección de datos, así como leer y escribir en archivos externos. También le permite agregar y agrupar datos para su análisis. Además, el proyecto incluye un tutorial completo de su API, con ejemplos de cada método, así como instrucciones para leer y escribir archivos CSV, que los científicos de datos suelen utilizar para preservar y distribuir datos tabulares.

**Apify** El proceso de extracción de datos se utilizó la herramienta de Apify, seguido del análisis de los datos resultantes utilizando GT/Pharo y DataFrame, una sólida herramienta de análisis y procesamiento de datos para lenguajes informáticos como Smalltalk. Evaluaremos varias características de los datos obtenidos, específicamente la dimensionalidad y la densidad, para obtener información valiosa sobre la calidad de los microdatos. La dimensionalidad se refiere a la cantidad de datos disponibles, mientras que la densidad indica cuán profundos son los datos, es decir, cuántos datos se agrupan. Nunca evaluamos la frecuencia de los trinos. Con estas definiciones claras, procederemos a analizar los datos recopilados. Nos centraremos en dos aspectos principales: la dimensionalidad y la

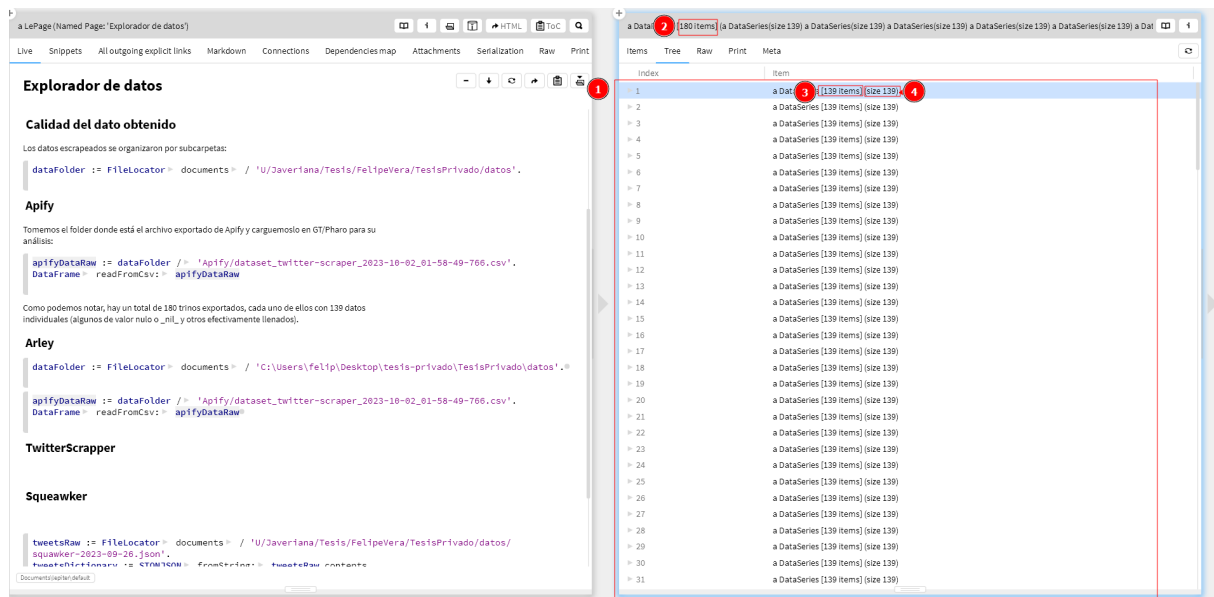
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

densidad de los tweets. La dimensionalidad se refiere a la cantidad de datos disponibles, mientras que la densidad se refiere a la profundidad de los datos, es decir, cuántos datos se agrupan. Estos análisis nos permitirán comprender mejor la calidad de los microdatos extraídos.

Tomemos el directorio donde se encuentra almacenado el archivo exportado de Apify y procedamos a cargarlo en GT/Pharo para su análisis

```
dataFolder := FileLocator documents /  
↳ 'C:\Users\felip\Desktop\tesis-privado\TesisPrivado\datos'.
```

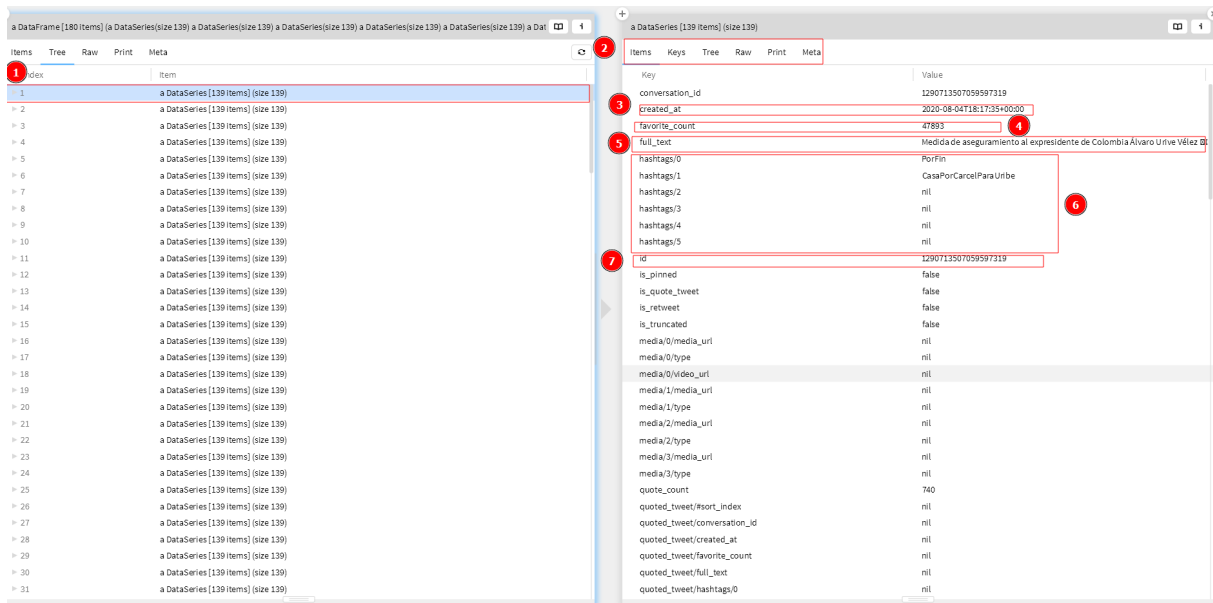
```
apifyDataRow := dataFolder /  
↳ 'Apify/dataset_twitter-scraper_2023-10-02_01-58-49-766.csv'.  
apifyData := DataFrame readFromCsv: apifyDataRow
```



**Figure 42:** Catalogo de datos Scrapeados

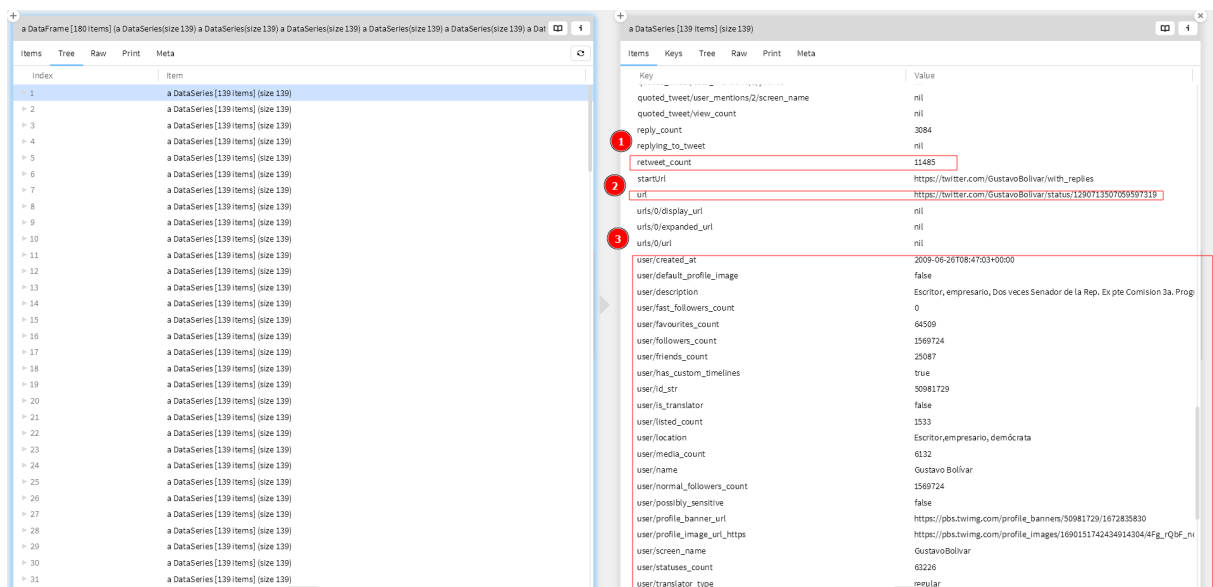
Como vemos se han recopilado 180 Tweets o trinos, mostrando cada Tweet sus 139 Items y 139 tamaños.

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



Cuando eliges un tweet, revelará todos sus datos descriptivos, incluidos:

1. Tweet de interés
2. Sección para búsquedas de datos de tweets
3. La fecha y hora de creación del tweet.
4. Conteo de “me gusta”
5. La cantidad y sustancia de los hashtags utilizados en un tweet.
6. ID de Twitter



1. Cantidad de Retweets



2. Url del tweet
3. Información de sobre el perfil de usuario de donde esta el tweet

```
apifyTweet := apifyData first.
```

Key	Value
conversation_id	1290713507059597319
created_at	2020-08-04T18:17:35+00:00
favorite_count	47893
full_text	Medida de aseguramiento al expresidente de Colombia Alvaro Uribe Vélez
hashtags/0	PorFin
hashtags/1	CasaPorCarcelParaUribe
hashtags/2	nil
hashtags/3	nil
hashtags/4	nil
hashtags/5	nil
id	1290713507059597319
is_pinned	false
is_quote_tweet	false
is_retweet	false
is_truncated	false
media/0/media_url	nil
media/0/type	nil
media/0/video_url	nil
media/1/media_url	nil
media/1/type	nil
media/2/media_url	nil
media/2/type	nil
media/3/media_url	nil
media/3/type	nil
quote_count	740
quoted_tweet/#sort_index	nil
quoted_tweet/conversation_id	nil
quoted_tweet/created_at	nil
quoted_tweet/favorite_count	nil
quoted_tweet/full_text	nil
quoted_tweet/hashtags/0	nil

**Figure 43:** Primer Tweet

Este comando especifica el primer tweet de la colección de tweets recopilados y muestra sus atributos, como por ejemplo:

1. La fecha y hora de creación del tweet.

2. Conteo de “me gusta”
3. Material textual
4. El número y nombre de los hashtags.
5. Microdatos y datos de tweets

Empezaremos por medir la dimensionalidad del trino, es decir cuántos datos de distintas características, o dimensiones, se encuentran en cada trino.

```
dimensions := (apifyTweet keys collect: [:key | (key splitOn: $/) first ]) asSet asOrderedCollection
```

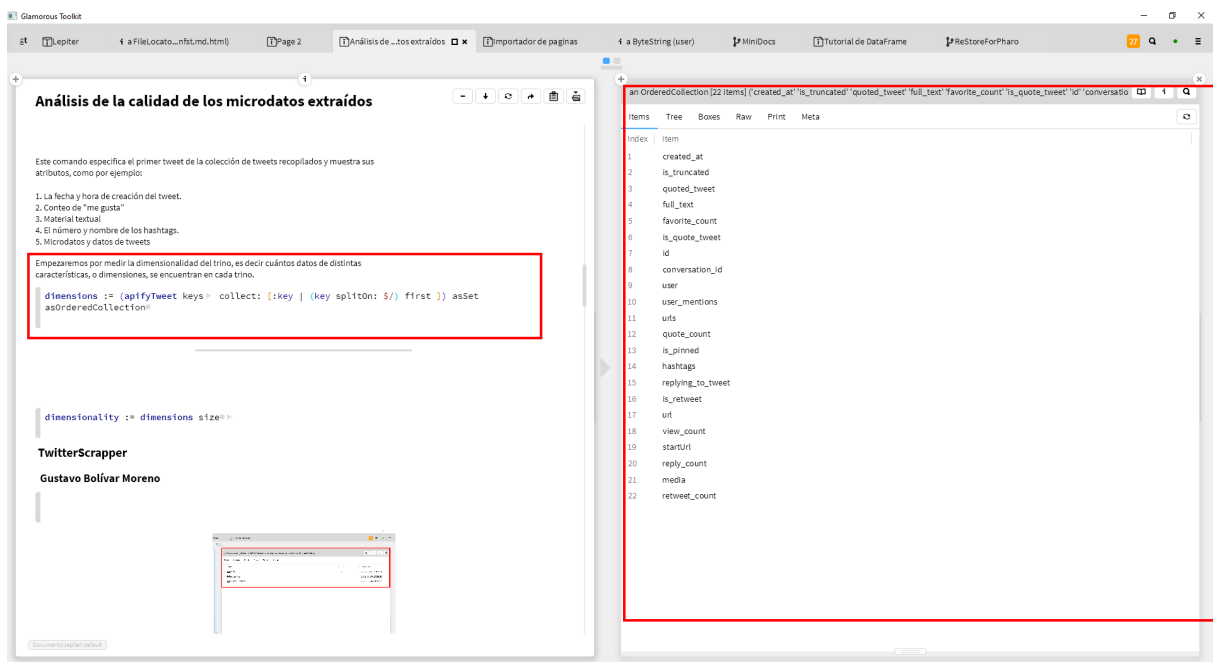
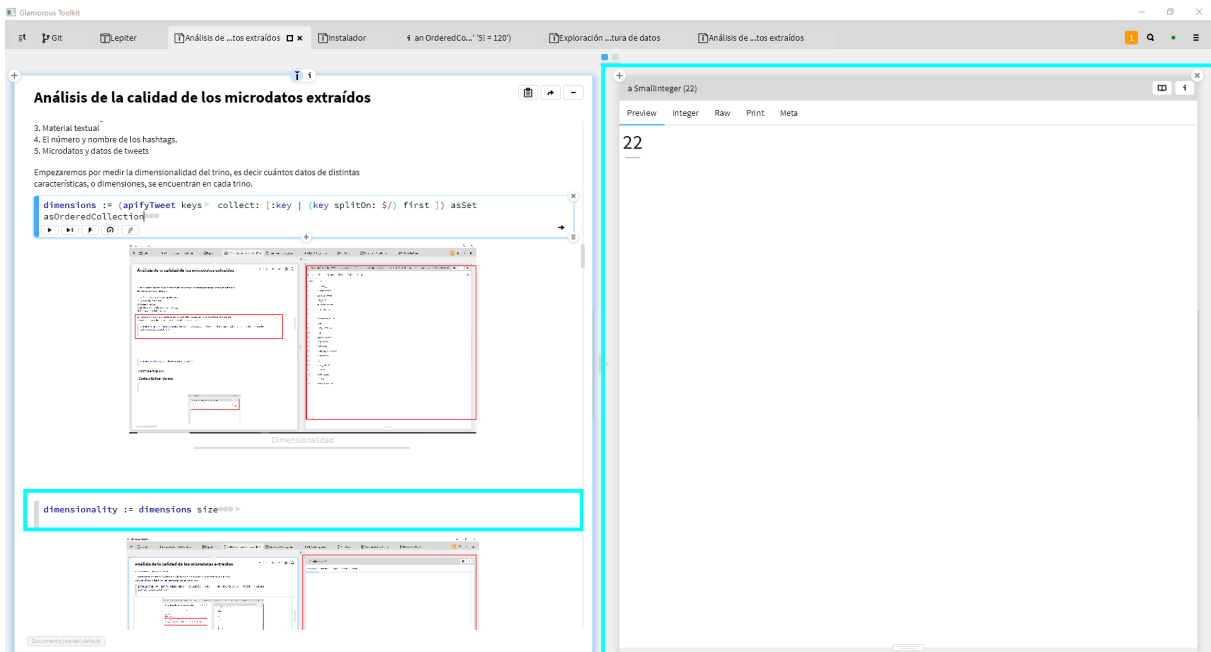


Figure 44: Dimensionalidad

```
dimensionality := dimensions size
```

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



**Figure 45:** Cantidad de características

Vamos a mirar la densidad de cada dato. Es decir, cuantos valores posibles hay agrupados en dicho dato, para todas las dimensiones antes halladas.

```
dataDensity := Dictionary new.
```

```
dimensions collect: [:dimension | | dataSet |
  dataSet := apifyTweet keys select: [:key | key beginsWith: dimension].
  dataDensity
    at: dimension
    put: dataSet size.
].
dataDensity
```

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

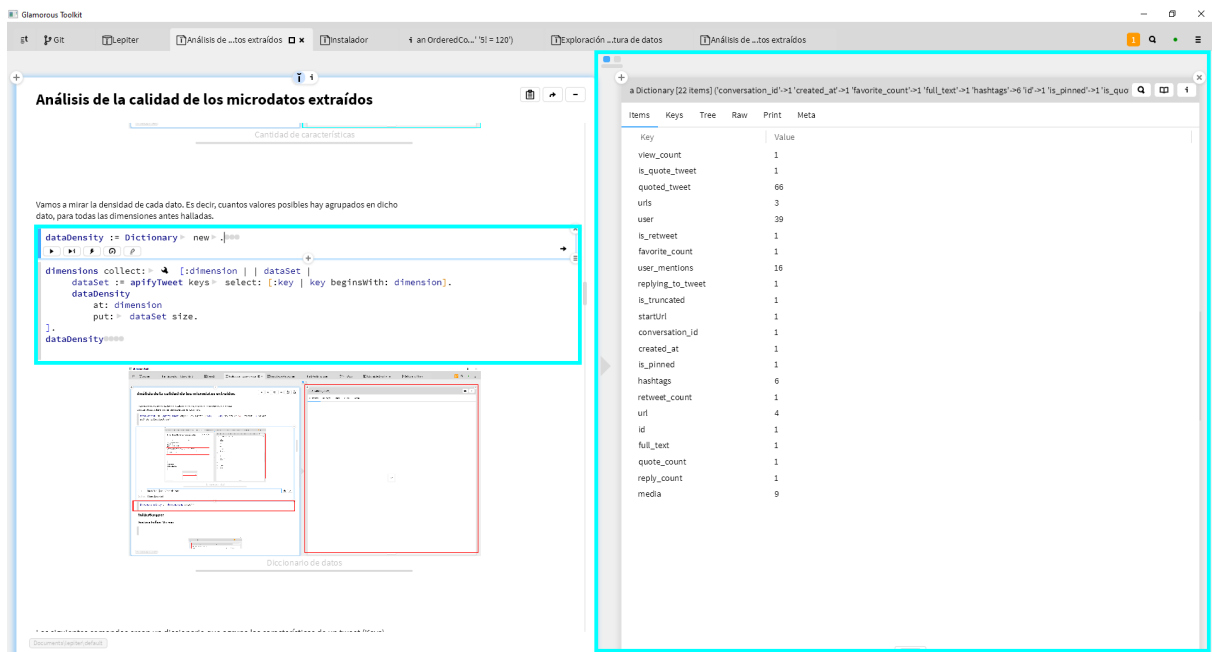


Figure 46: Diccionario de datos

Los siguientes comandos crean un diccionario que agrupa las características de un tweet (Keys), indica la cantidad de datos disponibles en cada uno. Es vital señalar que este proceso permite una clasificación sistemática de los elementos clave incluidos en los tweets, lo que facilita un mayor análisis y extracción de información significativa. Con estos comandos se creó un diccionario en donde clasifica las características principales del tweet “Keys” y muestra la cantidad de datos que se encuentran en dicha característica.

```
(apifyTweet keys select: [:key | key beginsWith: (dimensions at: 14)]) size
```

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

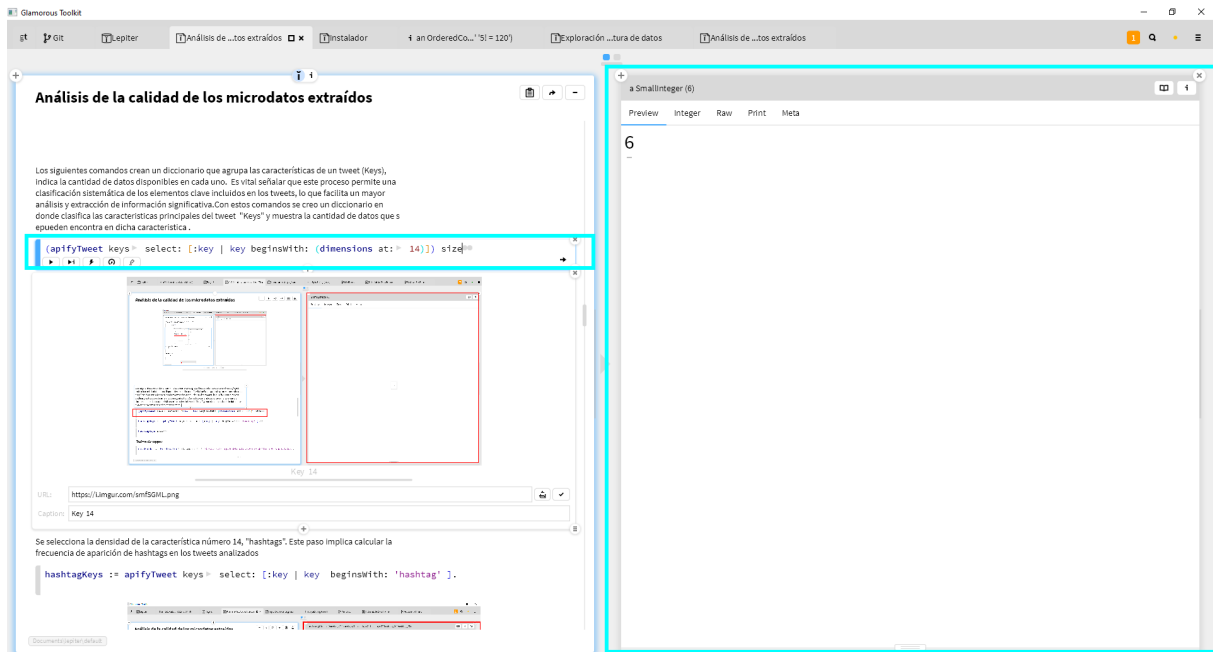


Figure 47: Key 14

Se selecciona la densidad de la característica número 14, “hashtags”. Este paso implica calcular la frecuencia de aparición de hashtags en los tweets analizados

```
hashtagKeys := apifyTweet keys select: [:key | key beginsWith: 'hashtag' ].
```

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

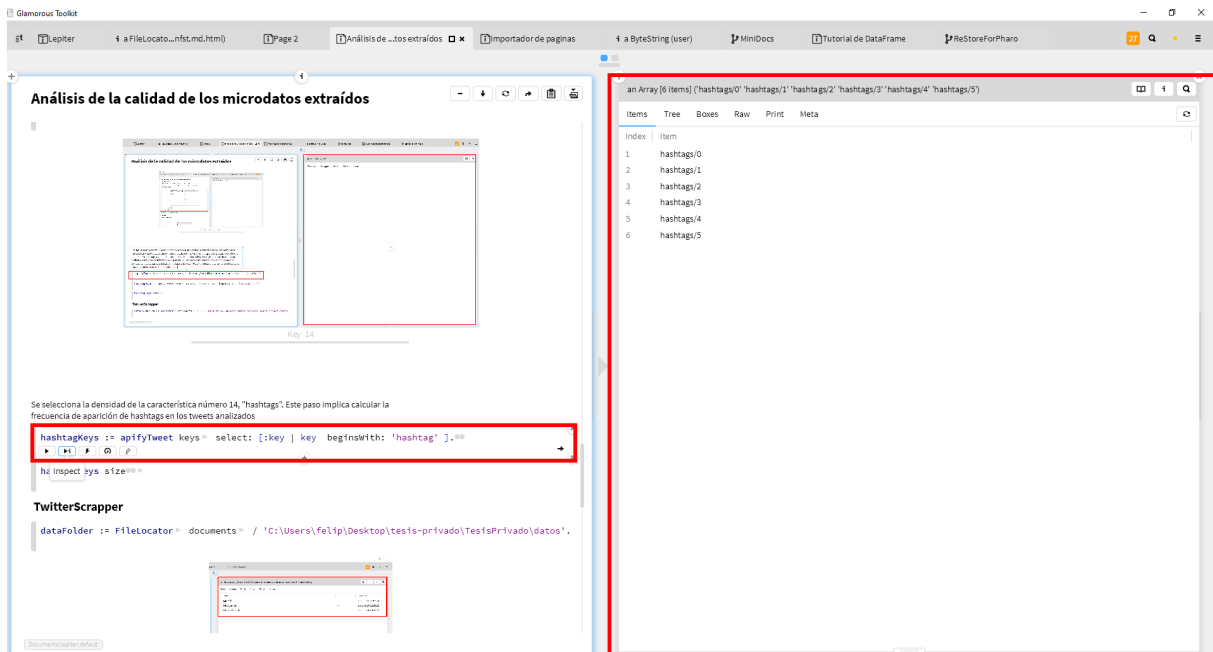


Figure 48: Key hashtags

hashtagKeys size

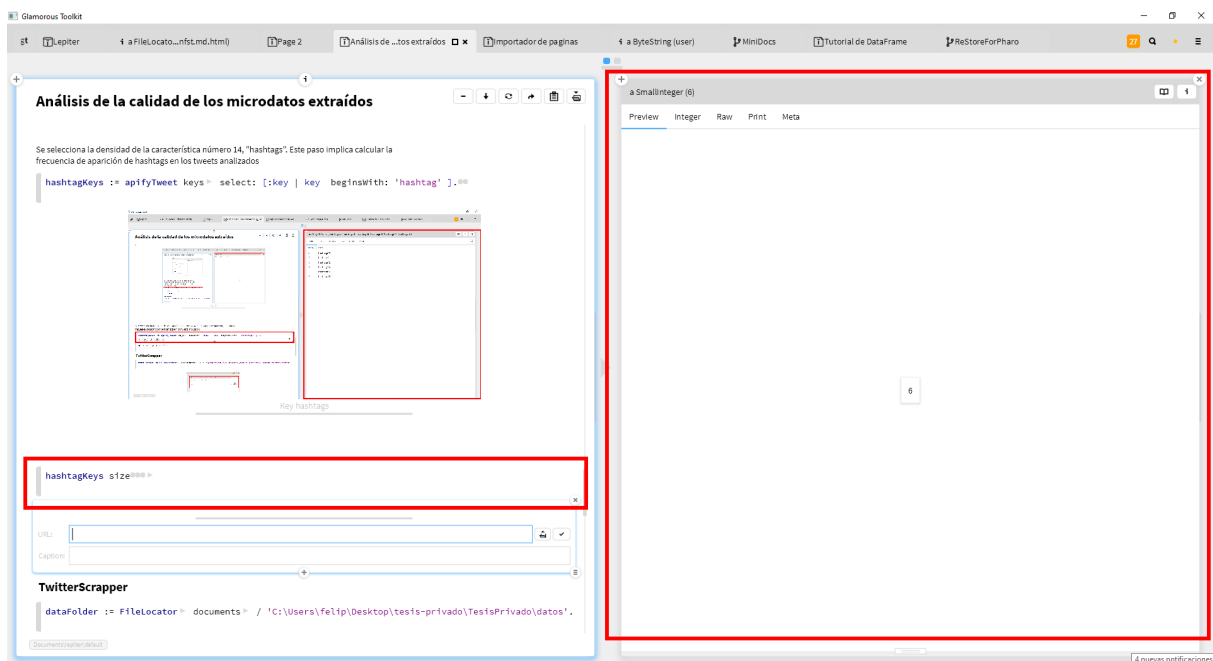
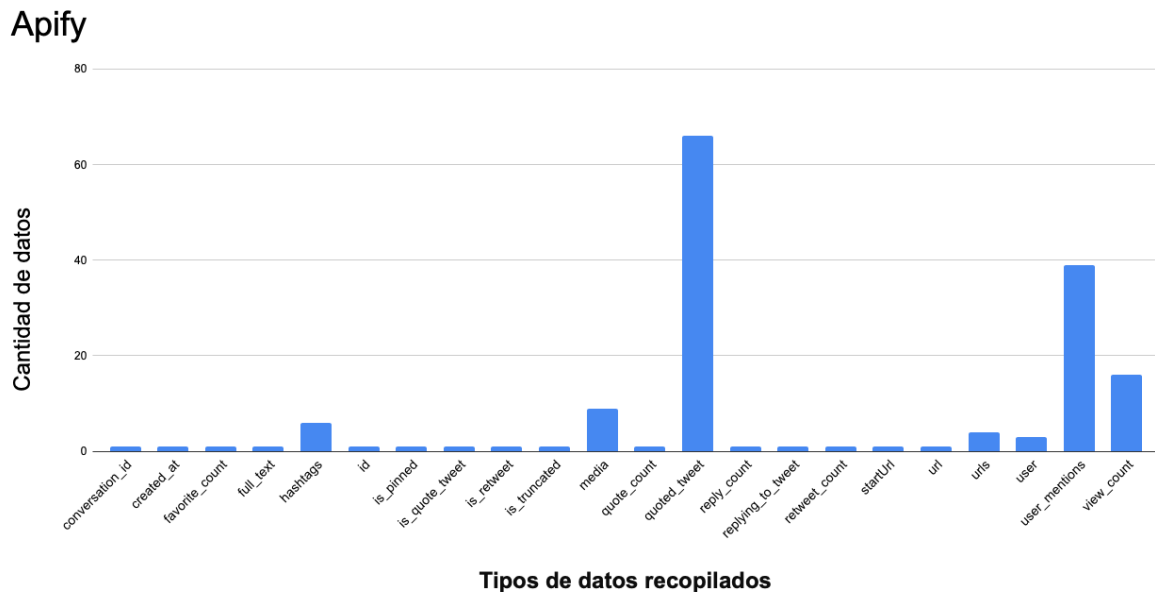


Figure 49: Size de hashtag

Muestra la densidad del Key “hashtag” que son 6.

La gráfica representa la cantidad de datos agrupados en función del tipo de dato, mostrando la distribución y clasificación de los datos recopilados. Es importante destacar que el eje horizontal que representa los diferentes tipos de datos recopilados, lo que proporciona una visualización más precisa de la variedad y cantidad de información obtenida durante el proceso de recolección de datos.



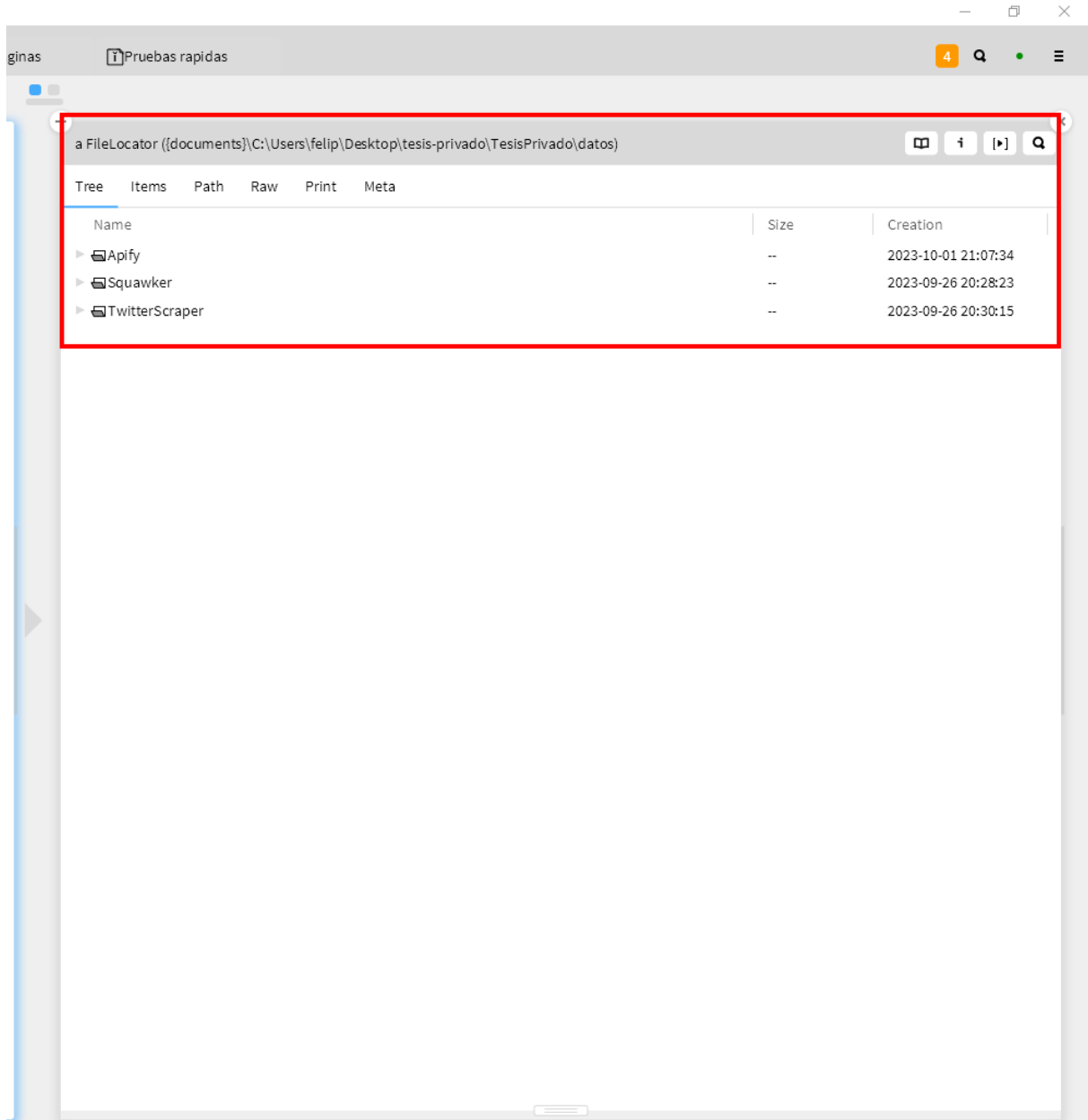
**Figure 50:** Gráfico representativo de los datos

En este estudio, utilizamos la herramienta Apify para recopilar datos de Twitter, que posteriormente analizamos con GT/Pharo. Investigamos la dimensionalidad de los tweets y utilizamos un histograma para evaluar la densidad de datos de cada característica. Basándonos en el histograma creado, destacamos el atributo “quoted\_tweet”, que tenía una densidad de datos significativamente mayor que las otras propiedades.

**TwitterScraper** En este estudio, utilizamos la aplicación TwitterScraper para examinar y recopilar datos completos de Twitter, como interacciones de usuarios, fechas de creación de tweets y contenido textual de tweets.

Centrándonos en las características incluidas en los tweets recopilados, reconocimos la relevancia del contenido del tweet, representado por la propiedad “Contenido”, que tenía una mayor densidad de datos que los otros atributos.

```
dataFolder := FileLocator documents /  
  ↪ 'C:\Users\felip\Desktop\tesis-privado\TesisPrivado\datos'.
```



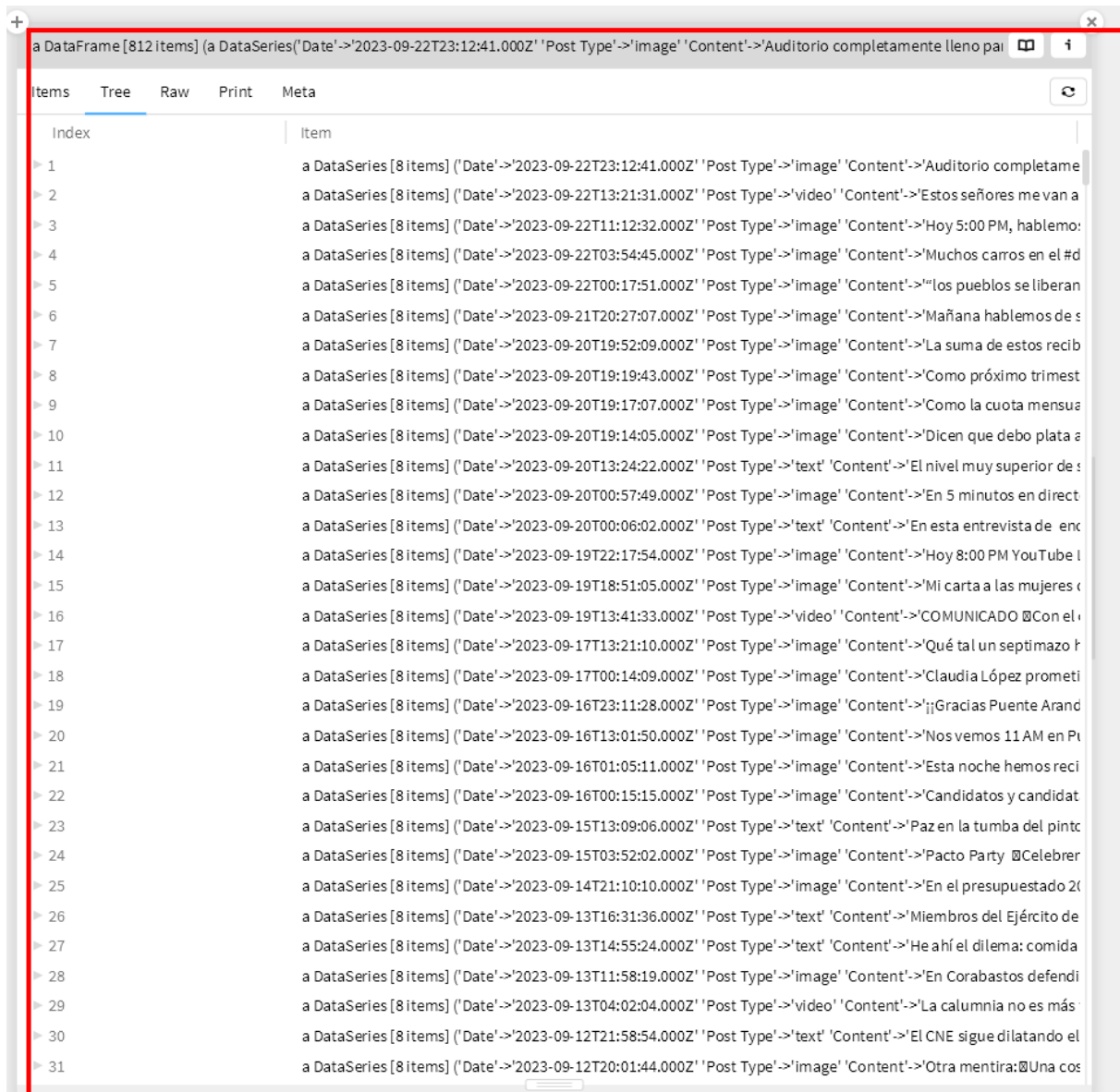
**Figure 51:** Origen de los datos Scrapeados

Este comando define los directorios en los que se almacenan los datos recuperados.

```
twitterScraperDataRow := dataFolder / 'TwitterScraper\GustavoBolivar.csv'.  
twitterScraperDataFrame := DataFrame readFromCsv: twitterScraperDataRow.
```



## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

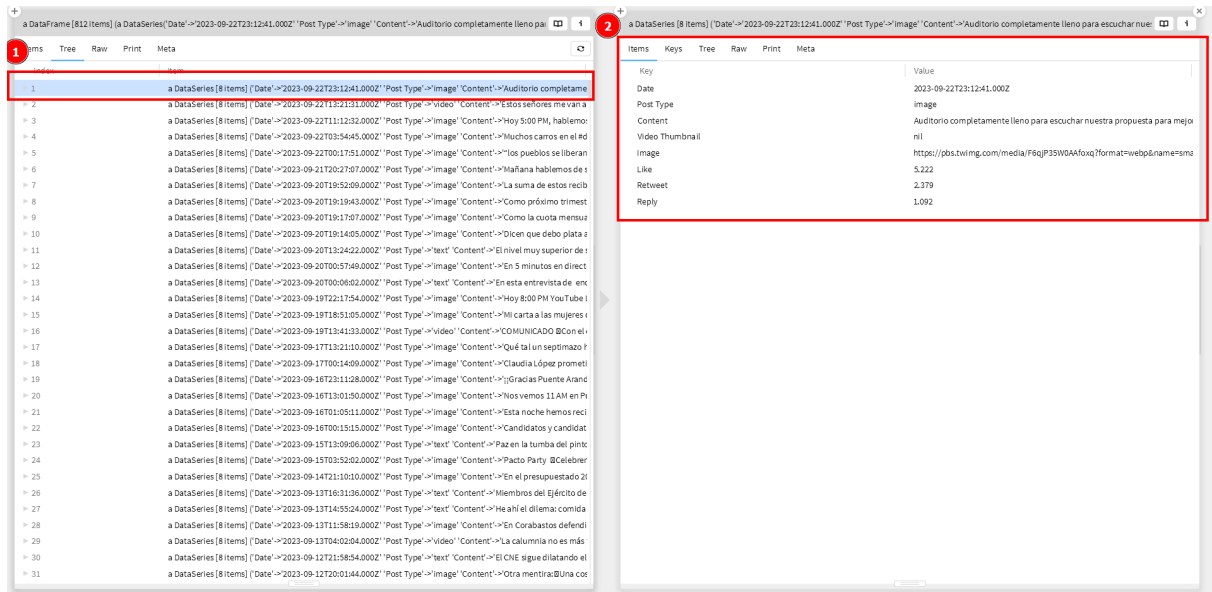


Index	Item
1	a DataService [8 items] ('Date'->'2023-09-22T23:12:41.000Z' 'Post Type'->'image' 'Content'->'Auditorio completame
2	a DataService [8 items] ('Date'->'2023-09-22T13:21:31.000Z' 'Post Type'->'video' 'Content'->'Estos señores me van a
3	a DataService [8 items] ('Date'->'2023-09-22T11:12:32.000Z' 'Post Type'->'image' 'Content'->'Hoy 5:00 PM, hablemo:
4	a DataService [8 items] ('Date'->'2023-09-22T03:54:45.000Z' 'Post Type'->'image' 'Content'->'Muchos carros en el #d
5	a DataService [8 items] ('Date'->'2023-09-22T00:17:51.000Z' 'Post Type'->'image' 'Content'->'"los pueblos se liberan
6	a DataService [8 items] ('Date'->'2023-09-21T20:27:07.000Z' 'Post Type'->'image' 'Content'->'Mañana hablemos de s
7	a DataService [8 items] ('Date'->'2023-09-20T19:52:09.000Z' 'Post Type'->'image' 'Content'->'La suma de estos recib
8	a DataService [8 items] ('Date'->'2023-09-20T19:19:43.000Z' 'Post Type'->'image' 'Content'->'Como próximo trimest
9	a DataService [8 items] ('Date'->'2023-09-20T19:17:07.000Z' 'Post Type'->'image' 'Content'->'Como la cuota mensua
10	a DataService [8 items] ('Date'->'2023-09-20T19:14:05.000Z' 'Post Type'->'image' 'Content'->'Dicen que debo plata z
11	a DataService [8 items] ('Date'->'2023-09-20T13:24:22.000Z' 'Post Type'->'text' 'Content'->'El nivel muy superior de s
12	a DataService [8 items] ('Date'->'2023-09-20T00:57:49.000Z' 'Post Type'->'image' 'Content'->'En 5 minutos en direct
13	a DataService [8 items] ('Date'->'2023-09-20T00:06:02.000Z' 'Post Type'->'text' 'Content'->'En esta entrevista de enc
14	a DataService [8 items] ('Date'->'2023-09-19T22:17:54.000Z' 'Post Type'->'image' 'Content'->'Hoy 8:00 PM YouTube l
15	a DataService [8 items] ('Date'->'2023-09-19T18:51:05.000Z' 'Post Type'->'image' 'Content'->'Mi carta a las mujeres c
16	a DataService [8 items] ('Date'->'2023-09-19T13:41:33.000Z' 'Post Type'->'video' 'Content'->'COMUNICADO @Con el
17	a DataService [8 items] ('Date'->'2023-09-17T13:21:10.000Z' 'Post Type'->'image' 'Content'->'Qué tal un septimazo f
18	a DataService [8 items] ('Date'->'2023-09-17T00:14:09.000Z' 'Post Type'->'image' 'Content'->'Claudia López prometi
19	a DataService [8 items] ('Date'->'2023-09-16T23:11:28.000Z' 'Post Type'->'image' 'Content'->'¡¡Gracias Puente Arand
20	a DataService [8 items] ('Date'->'2023-09-16T13:01:50.000Z' 'Post Type'->'image' 'Content'->'Nos vemos 11 AM en Pt
21	a DataService [8 items] ('Date'->'2023-09-16T01:05:11.000Z' 'Post Type'->'image' 'Content'->'Esta noche hemos reci
22	a DataService [8 items] ('Date'->'2023-09-16T00:15:15.000Z' 'Post Type'->'image' 'Content'->'Candidatos y candidat
23	a DataService [8 items] ('Date'->'2023-09-15T13:09:06.000Z' 'Post Type'->'text' 'Content'->'Paz en la tumba del pintc
24	a DataService [8 items] ('Date'->'2023-09-15T03:52:02.000Z' 'Post Type'->'image' 'Content'->'Pacto Party @Celebrer
25	a DataService [8 items] ('Date'->'2023-09-14T21:10:10.000Z' 'Post Type'->'image' 'Content'->'En el presupuestado 2(
26	a DataService [8 items] ('Date'->'2023-09-13T16:31:36.000Z' 'Post Type'->'text' 'Content'->'Miembros del Ejército de
27	a DataService [8 items] ('Date'->'2023-09-13T14:55:24.000Z' 'Post Type'->'text' 'Content'->'He ahí el dilema: comida
28	a DataService [8 items] ('Date'->'2023-09-13T11:58:19.000Z' 'Post Type'->'image' 'Content'->'En Corabastos defendi
29	a DataService [8 items] ('Date'->'2023-09-13T04:02:04.000Z' 'Post Type'->'video' 'Content'->'La calumnia no es más
30	a DataService [8 items] ('Date'->'2023-09-12T21:58:54.000Z' 'Post Type'->'text' 'Content'->'El CNE sigue dilatando el
31	a DataService [8 items] ('Date'->'2023-09-12T20:01:44.000Z' 'Post Type'->'image' 'Content'->'Otra mentira: @Una cos

**Figure 52:** Listado de datos Scrapeados

Este comando mostrará todos los tweets recopilados del candidato (Gustavo Bolivar) a la alcaldía de Bogotá.

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

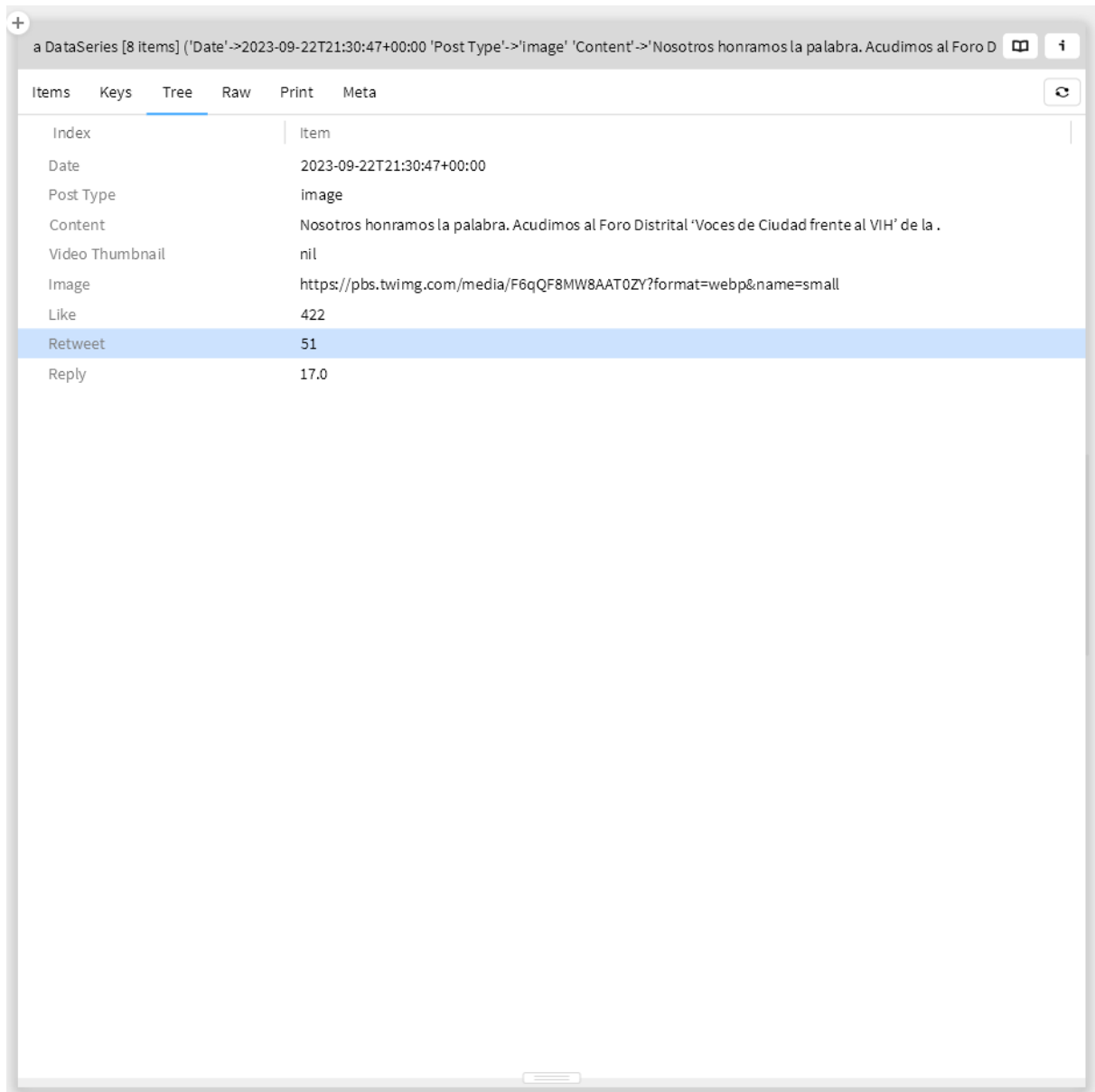


**Figure 53:** Características del tweet seleccionado

Este comando también se encarga de recopilar y mostrar datos precisos sobre un tweet seleccionado, creando una nueva pestaña con datos importantes como la fecha de creación, la cantidad de me gusta, retweets y comentarios, el tipo de tweet y su texto. Al organizar la información principal de un tweet en un marco fácil de analizar, este enfoque agiliza la investigación.

```
twitterScrapertweet := twitterScrapertweet first.
```

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



The screenshot shows a data viewer interface with a table of tweet metadata. The 'Retweet' row is highlighted in blue. The table contains the following data:

Index	Item
Date	2023-09-22T21:30:47+00:00
Post Type	image
Content	Nosotros honramos la palabra. Acudimos al Foro Distrital 'Voces de Ciudad frente al VIH' de la .
Video Thumbnail	nil
Image	<a href="https://pbs.twimg.com/media/F6qQF8MW8AAT0ZY?format=webp&amp;name=small">https://pbs.twimg.com/media/F6qQF8MW8AAT0ZY?format=webp&amp;name=small</a>
Like	422
Retweet	51
Reply	17.0

**Figure 54:** Primer Tweet seleccionado

```
dimensions := (twitterScrapperTweet keys collect: [:key | (key splitOn: $/)
  ↪ first ]) asSet asOrderedCollection
```

El comando recopila todas las características “Keys”

```
dataDensity := dimensions size
```

```
dataDensity := Dictionary new.
```

```
dimensions collect: [:dimension | | dataSet |
```

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

```
dataSet := twitterScrapperTweet keys select: [:key | key beginsWith:  
↪ dimension].  
dataDensity  
  at: dimension  
  put: dataSet size.  
].  
dataDensity
```

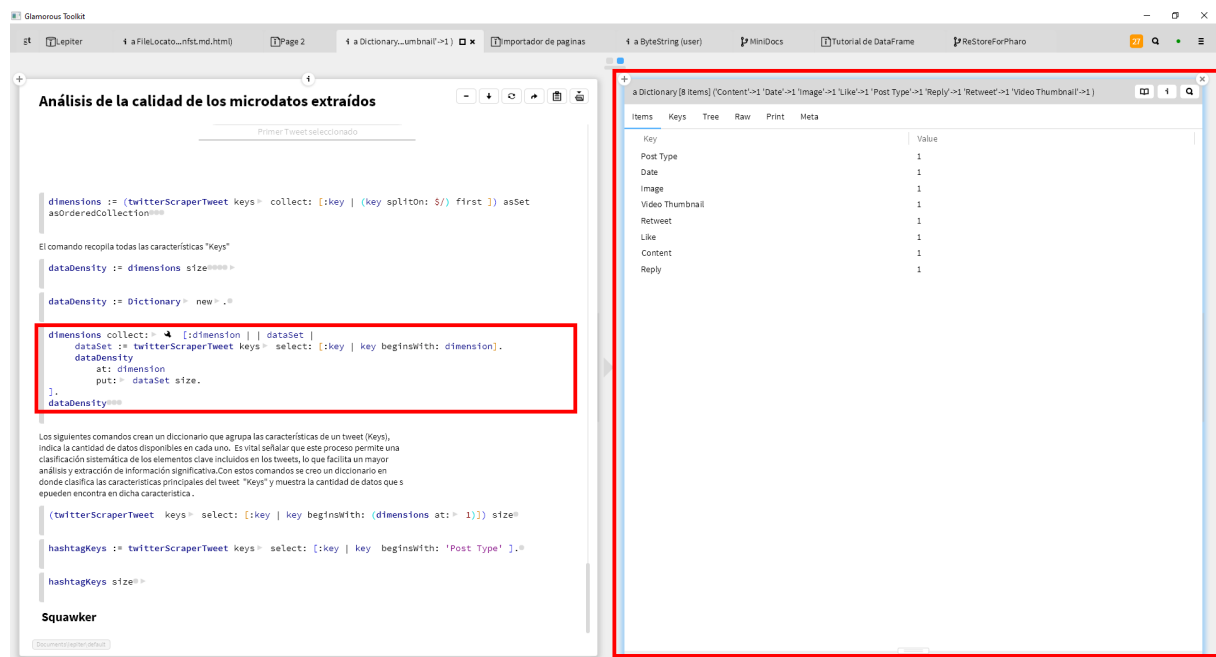
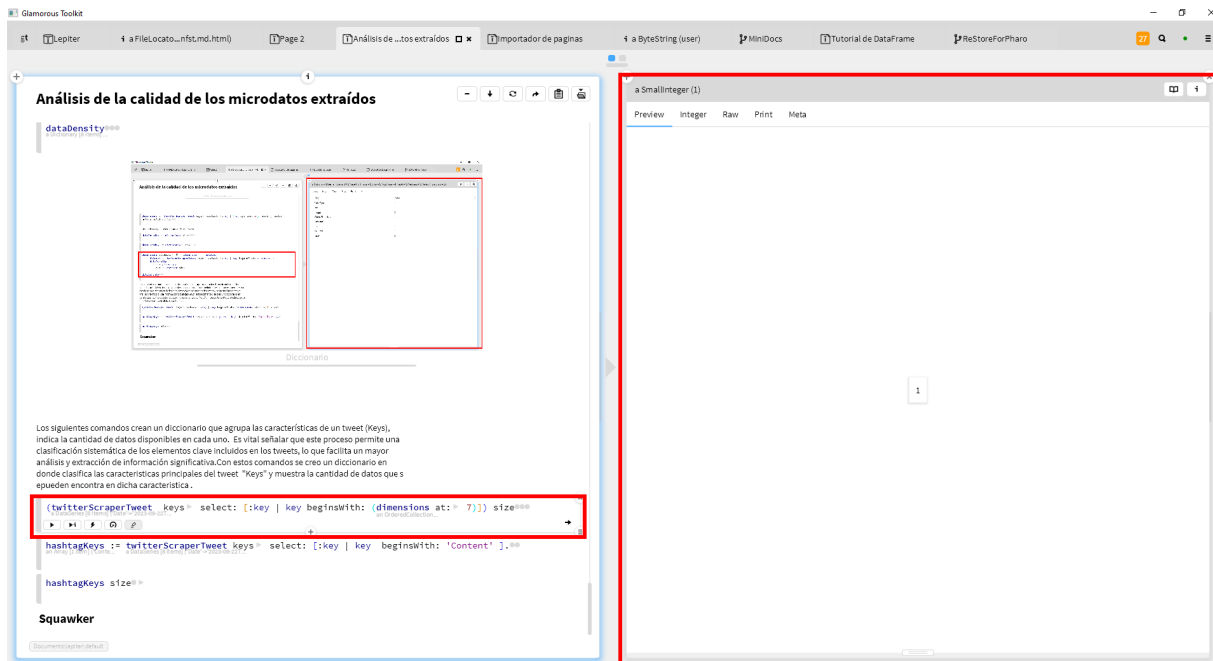


Figure 55: Diccionario

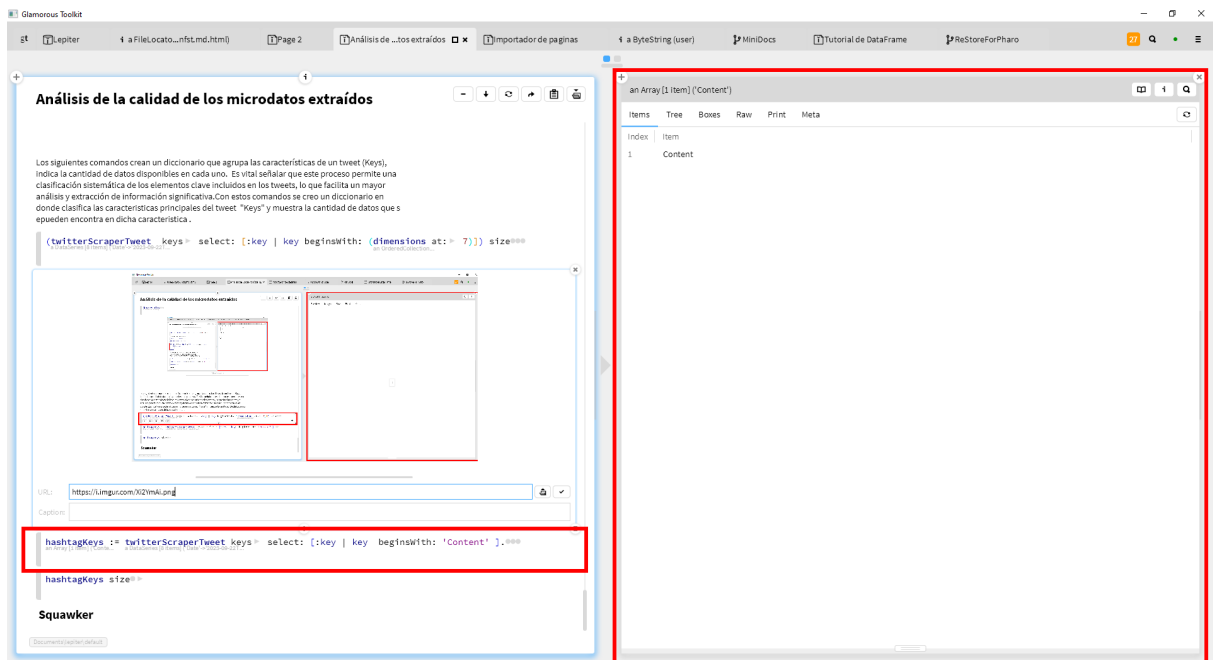
Los siguientes comandos crean un diccionario que agrupa las características de un tweet (Keys), indica la cantidad de datos disponibles en cada uno. Es vital señalar que este proceso permite una clasificación sistemática de los elementos clave incluidos en los tweets, lo que facilita un mayor análisis y extracción de información significativa. Con estos comandos se creó un diccionario en donde clasifica las características principales del tweet “Keys” y muestra la cantidad de datos que se encuentran en dicha característica.

```
(twitterScrapperTweet keys select: [:key | key beginsWith: (dimensions at:  
↪ 7)]) size
```

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

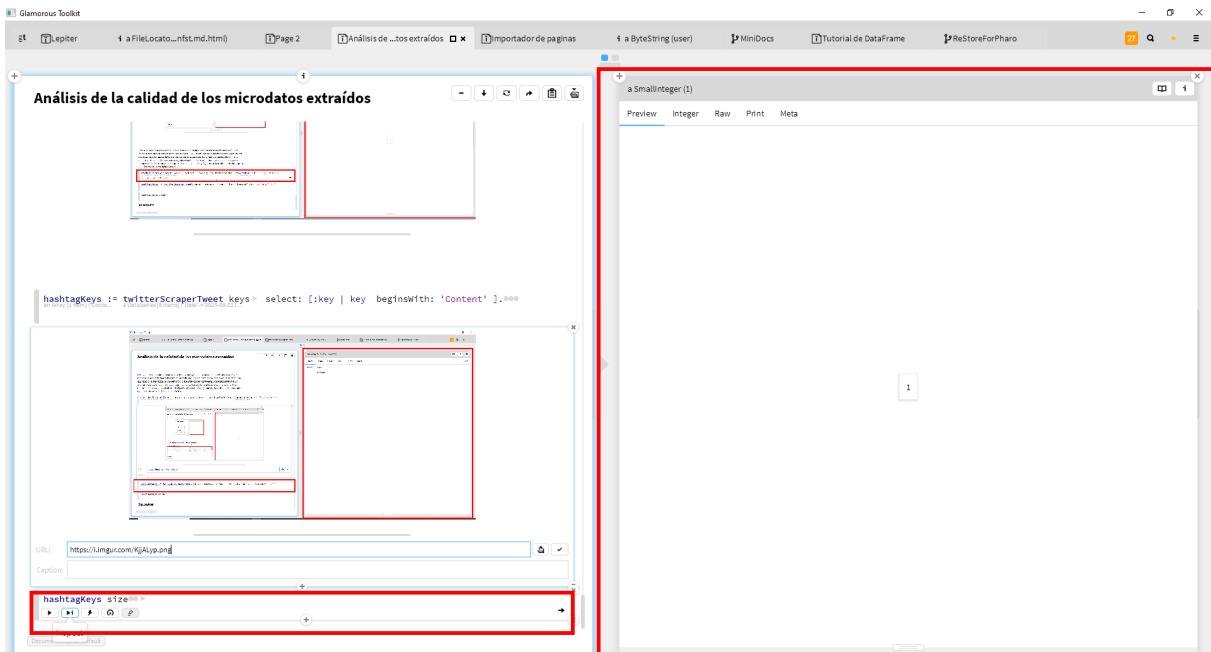


```
hashtagKeys := twitterScraPerTweet keys select: [:key | key beginsWith: 'Content' ].
```



```
hashtagKeys size
```

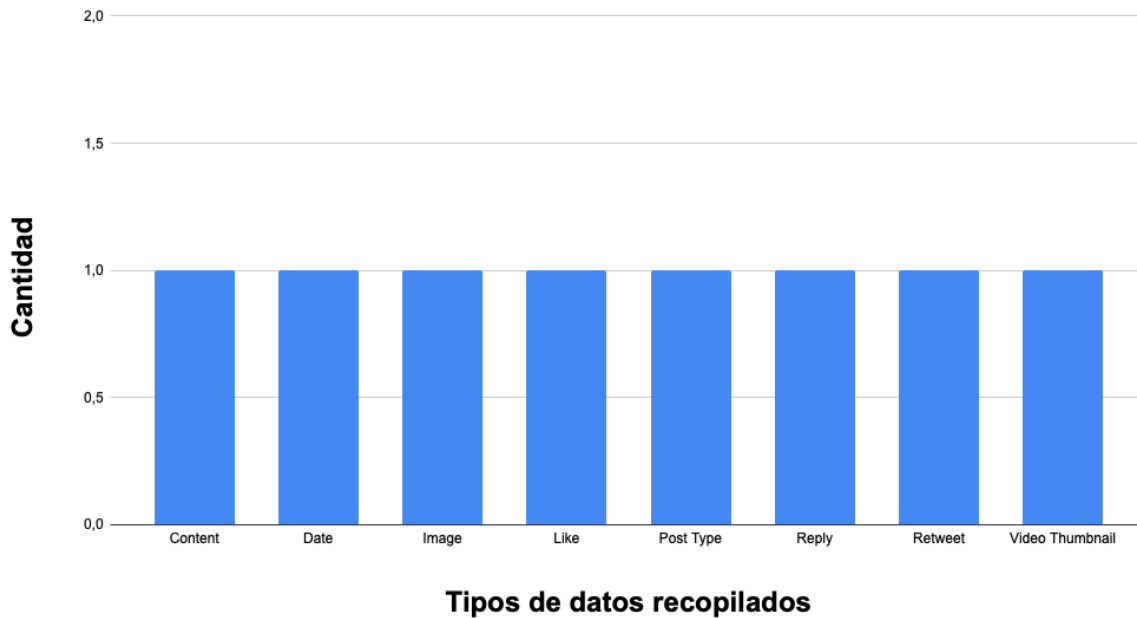
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



Muestra la densidad del Key “content” que es 1.

La gráfica representa la distribución de datos recopilados por Twitterscraper en función de diferentes categorías o tipos de información. En este contexto, la gráfica muestra la cantidad de datos agrupados según su clasificación, lo que permite visualizar de manera clara la diversidad de información obtenida a través de Twitterscraper. Es importante señalar que el eje horizontal refleja las distintas categorías de datos recopilados, brindando una representación detallada de la variedad de información extraída durante el proceso de scraping de Twitter.

## TwitterScraper



**Figure 56:** Gráfico representativo de los datos

En este estudio, utilizamos la herramienta TwitterScraper para evaluar los datos recopilados de los tweets relacionados con la campaña electoral de los candidatos para la alcaldía de Bogotá. Recopilamos datos detallados de Twitter, como interacciones de los usuarios, fechas de creación de tweets y contenido textual.

A pesar de nuestras expectativas, ningún atributo destacó cuando se evaluó mediante un histograma de las 8 Keys o Items recopilados. Esto ocurrió porque cada elemento incluía solo un dato, lo que resultó en una distribución equitativa de la densidad de datos entre las características

**Squawker** En este estudio, utilizamos la herramienta Squawker para producir los datos almacenados en archivos JSON. Estos archivos incluyen amplia información sobre los tweets recopilados. Sin embargo, para realizar un análisis significativo, la estructura de datos tuvo que convertirse de un formato de árbol a un formato plano.

Para ello, empleamos herramientas como GT/Pharo para analizar datos de archivos JSON. Con esta herramienta se pudo empezar a visualizar los datos arbóreos extraídos de Squawker y efectivamente constatar en una aproximación informal, que se trata de uno de los scrappers de mayor calidad de los acá estudiandos. Sin embargo, los límites de tiempo y conocimientos sólo permitieron adelantar

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

cómo importar datos y empezar con algunas visualizaciones. Un estudio posterior de dicha fuente de datos se deja para las recomendaciones (como se muestra en la sección respectiva).

reemplazar

```
dataFolder := FileLocator documents /  
↳ 'C:\Users\felip\Desktop\tesis-privado\TesisPrivado\datos\Squawker'
```

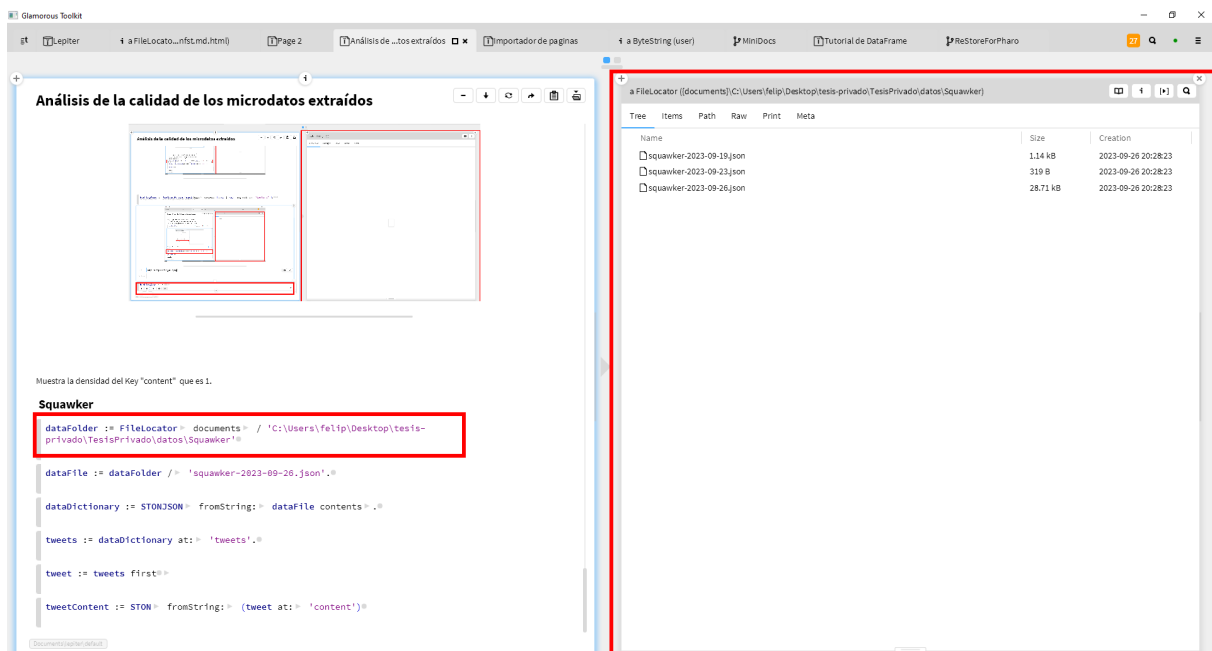


Figure 57: Ubicacion

```
dataFile := dataFolder / 'squawker-2023-09-26.json'.
```



# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

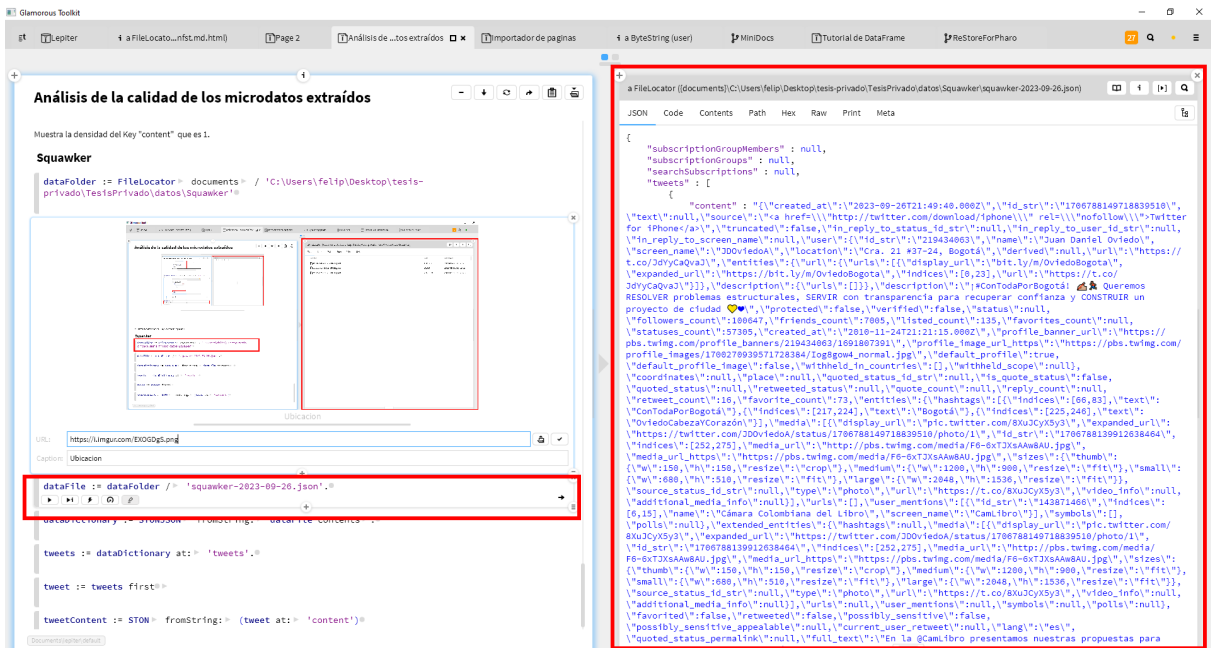
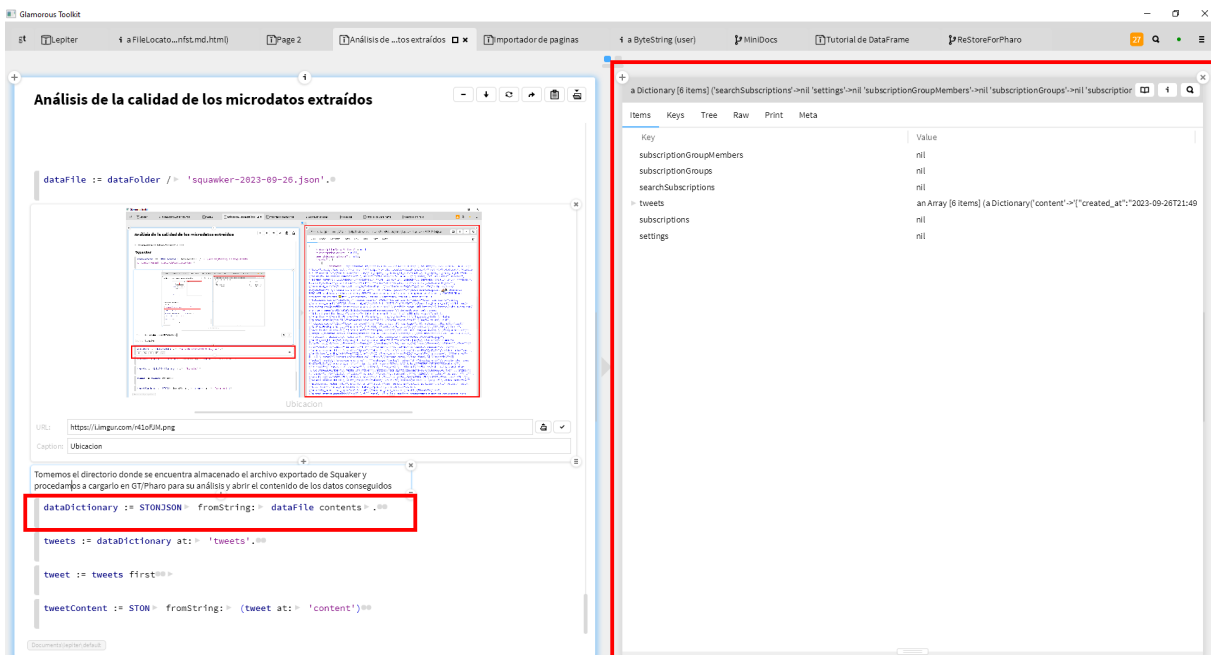


Figure 58: Ubicacion

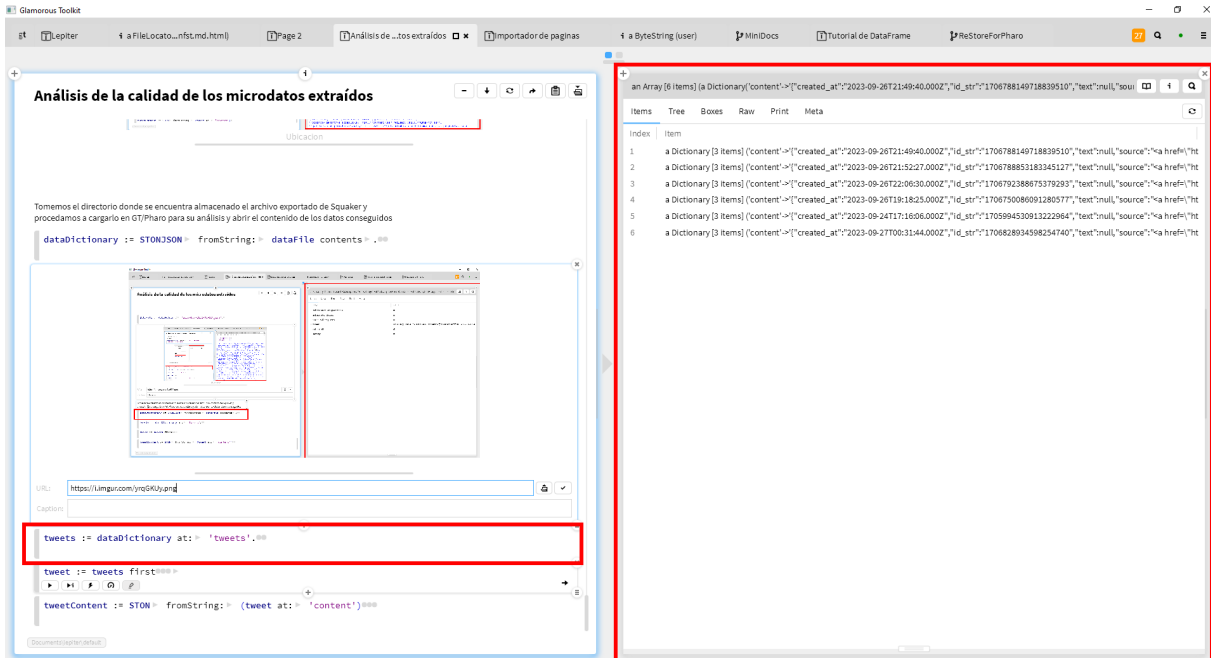
Tomemos el directorio donde se encuentra almacenado el archivo exportado de Squaker y procedamos a cargarlo en GT/Pharo para su análisis y abrir el contenido de los datos conseguidos

```
dataDictionary := STONJSON fromString: dataFile contents.
```

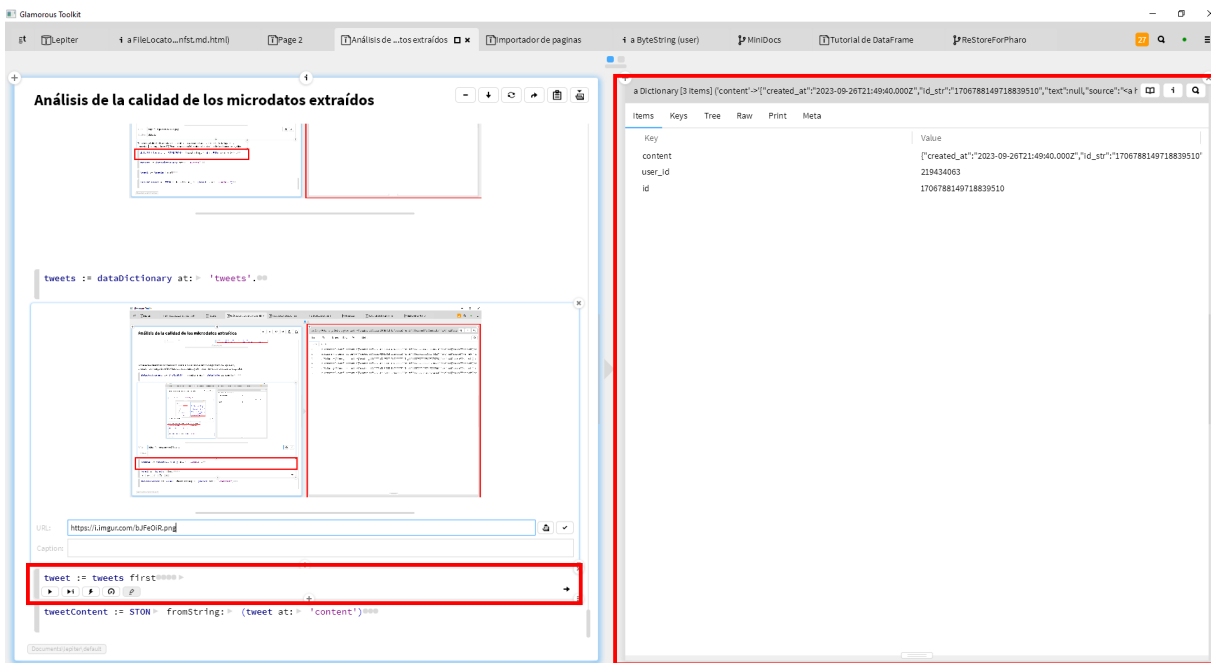


# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

```
tweets := dataDictionary at: 'tweets'.
```

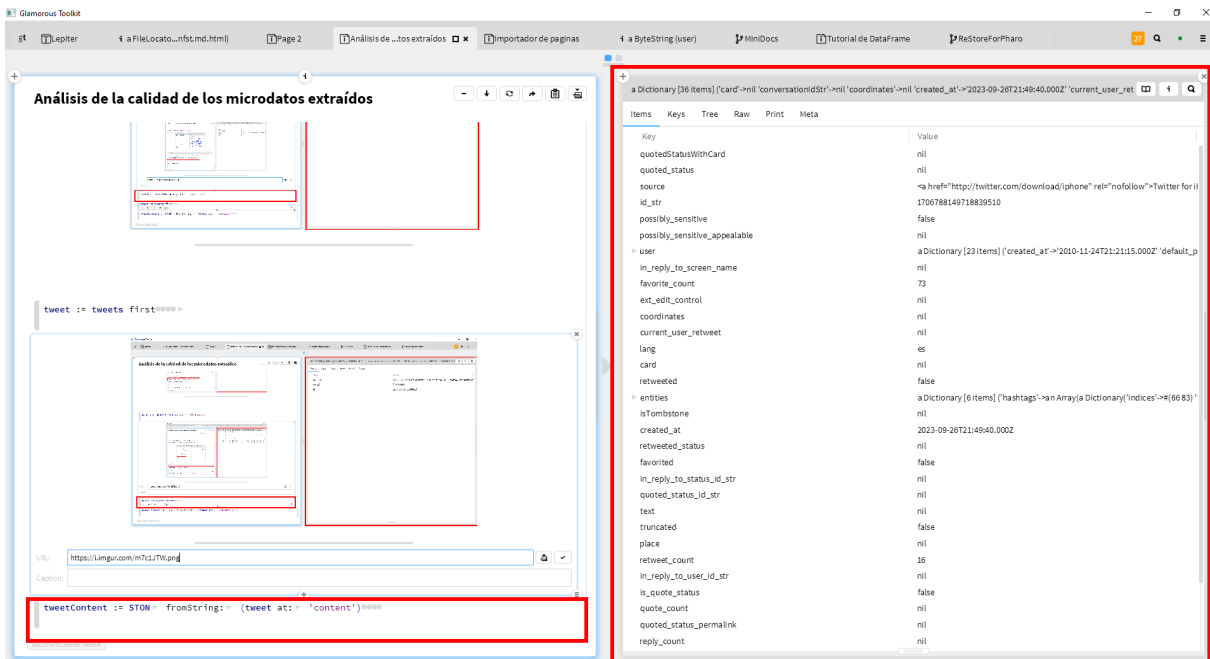


```
tweet := tweets first
```



```
tweetContent := STON fromString: (tweet at: 'content')
```

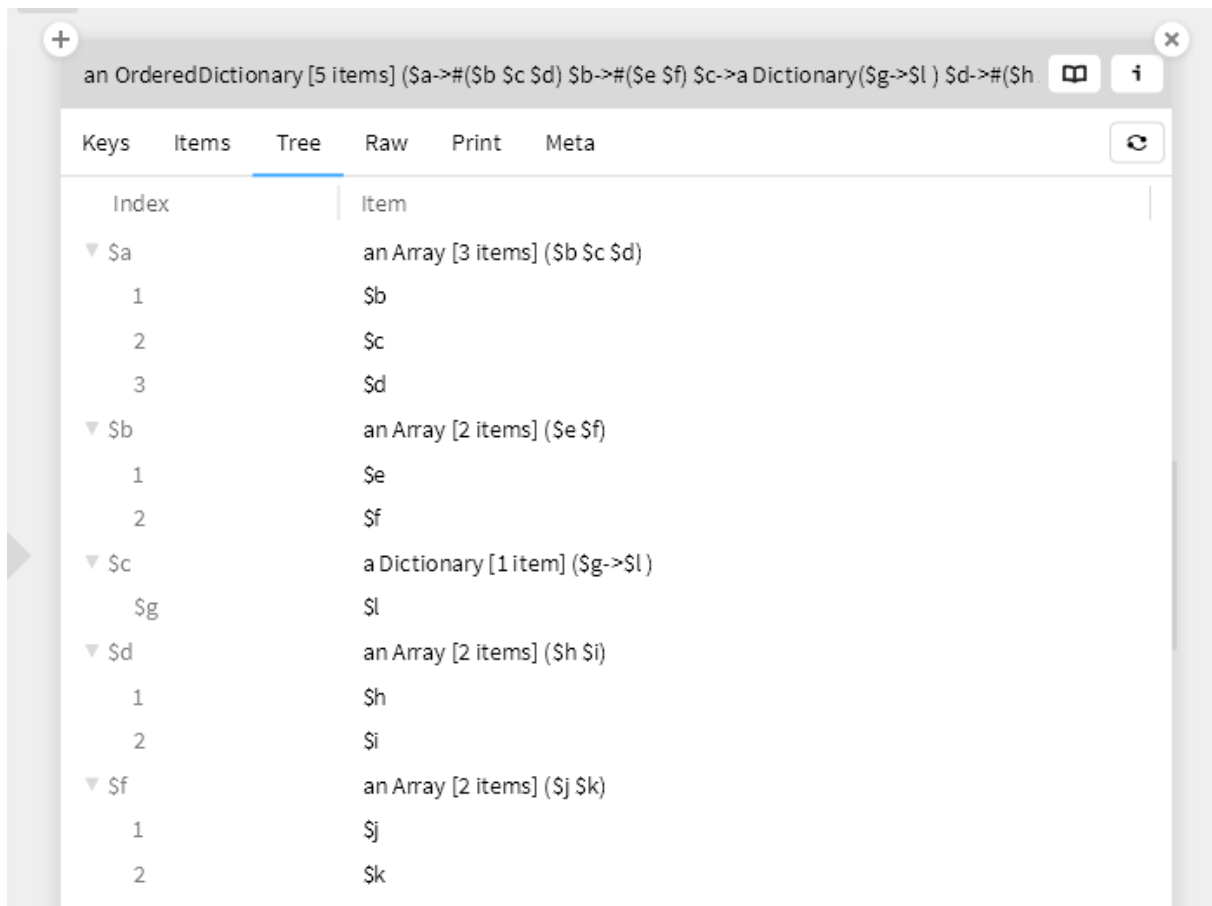
## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



Se especifica la ubicación del tweet que se desea analizar y se especifica en la característica “content” para visualizar los datos encontrados. Esta acción implica identificar un tweet específico dentro de un conjunto de datos y extraer información relevante contenida en la propiedad “content”.

Como se ha dicho, la información en Squakwer está organizada en forma de árbol. Por tanto, para su análisis, es importante poder declarar árboles y graficarlos. En esta sección iniciaremos con definiendo y visualizando un árbol sencillo de datos, para detectar los problemas particulares en ese tipo de estructuras de datos y luego sí pasaremos a árboles de datos más complejos como los que genera Squawker de su extracción de datos de Twitter. Implementemos los contenidos de un árbol como un diccionario en Pharo:

```
tree := OrderedDictionary new
  at: $a put: #( $b $c $d );
  at: $b put: #( $e $f );
  at: $c put: #($l);
  at: $d put: #( $h $i );
  at: $f put: #( $j $k);
  yourself
```

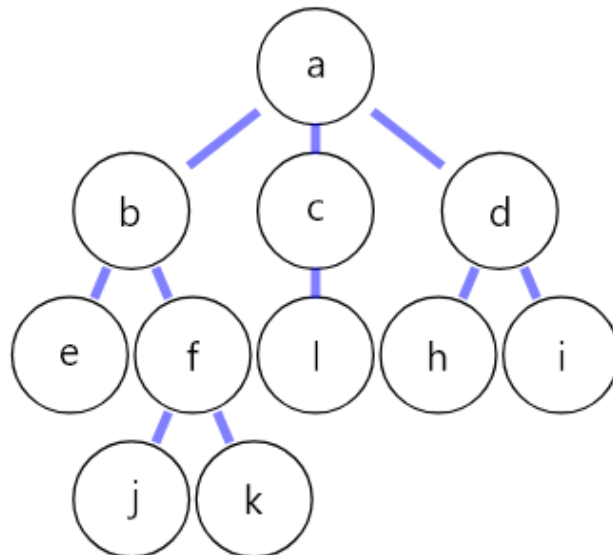


**Figure 59:** Diccionario sencillo como árbol de datos

Esta parte dibuja el árbol de datos anterior:

```
view := GtMondrian new.
view nodes
  stencil: [ :x |
    BElement new
      border: (BlBorder paint: Color black);
      geometry: BlEllipseGeometry new;
      layout: (BlLinearLayout new alignCenter);
      addChild: (BlTextElement text: (x asRopedText fontSize: 17)) ];
  with: (tree flatCollectAsSet: #yourself) , tree keys.
view edges
  stencil: [ :x | BlLineElement new border: (BlBorder paint: (Color blue
↪ alpha: 0.5) width: 4) ];
  connect: tree associations from: #key toAll: #value.
view layout tree.
```

view



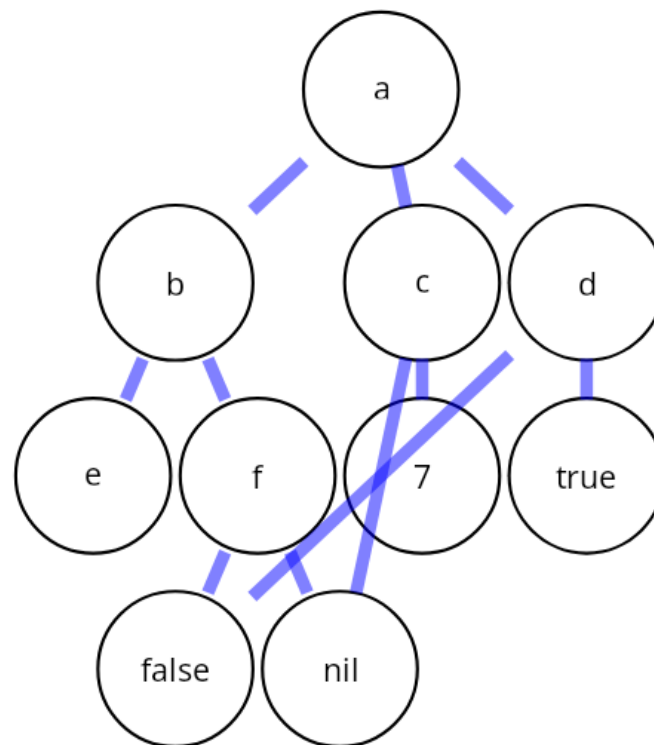
**Figure 60:** Árbol de datos a partir de un diccionario

Sin embargo, en las exploraciones se pudo apreciar que, cuando redefinimos el árbol de datos, de manera que tenga datos nulos, el árbol deja de funcionar, debido a que el dato nulo (o nil en Smalltalk) corresponde a una palabra reservada, que expresa un valor único. Por el contrario, los valores de verdadero (true) y falso (false) son instancias que sí pueden repetirse. Por ejemplo, si redefinimos el árbol de datos como:

```
tree := OrderedDictionary new
  at: $a put: #( $b $c $d );
  at: $b put: #( $e $f);
  at: $c put: #(7 nil);
  at: $d put: #( true false );
  at: $f put: #( false nil);
  at:$e put: #( );
  yourself
```

Para invocarlo de modo más sencillo, el tutor de la tesis extendió los métodos del diccionario de manera que se puede realizar una gráfica similar a la del árbol anterior, con los nuevos datos, de la siguiente manera:

```
tree treeView
```

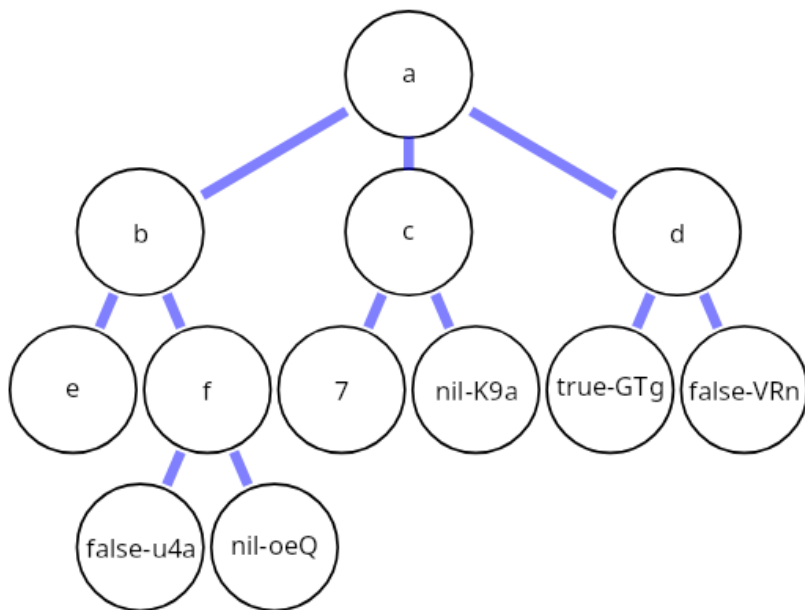


**Figure 61:** Árbol de datos con la estructura dañada, debido a que los valores nulos y booleanos apuntan a un sólo valor.

Se necesita reemplazar cada valor nulo o booleano por un identificador que nos permita preservar la estructura de árbol. Pues, como se ve en la gráfica anterior, el árbol se convierte en un grafo cíclico, en la medida que muchos nodos apuntan a un único valor nulo o booleano. Se usará el siguiente método, también creado por el tutor de la tesis en el marco de la misma y colocado por él dentro del paquete MiniDocs, precisamente para resolver el problema de crear identificadores únicos para los valores nulos, mostrando así también la idea de herramientas amoldables que se pueden adecuar al problema:

```
(tree replaceWithUniqueNilsAndBooleansStartingAt: 1) treeView
```

Lo anterior produce esta gráfica:



**Figure 62:** Árbol de datos con valores nulos y booleanos únicos.

Es posible apreciar que este método permite desambiguar los valores únicos de los booleanos y el valor nulo. Con lo cual, la gráfica del árbol puede ser rehecha de manera correcta, como se muestra en la figura previa. La siguiente cosa que hay que hacer es colocar los datos extraídos en Squawker como diccionario en tree y graficarlos:

```
dataRaw := FileLocator aliases at: 'squawkerJSON'.
data := STONJSON fromString: dataRaw contents.
tweet := (data at: 'tweets') first at: 'content'.
tree := STONJSON fromString: tweet contents.
```

```
tree replaceWithUniqueNilsAndBooleansStartingAt: 1
```

```
tree := STONJSON fromString: (FileLocator temp / 'dataTree.ston') contents
```

```
tree treeView
```

Sin embargo, al intentar graficar el árbol con los datos reales se obtuvo el siguiente error:

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

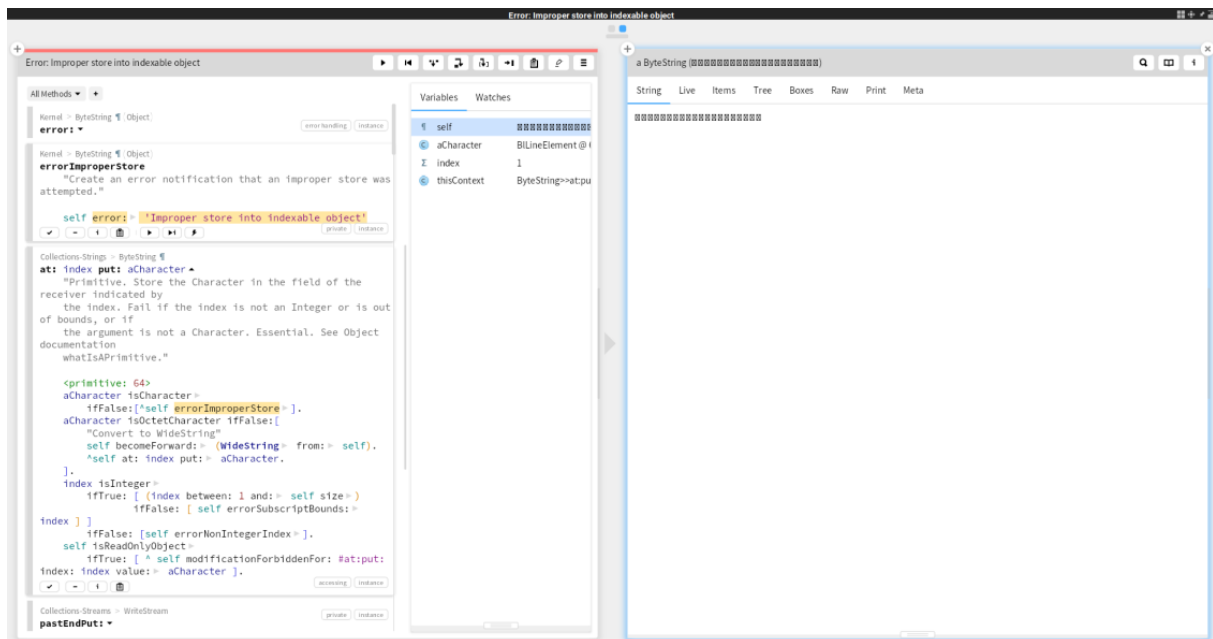
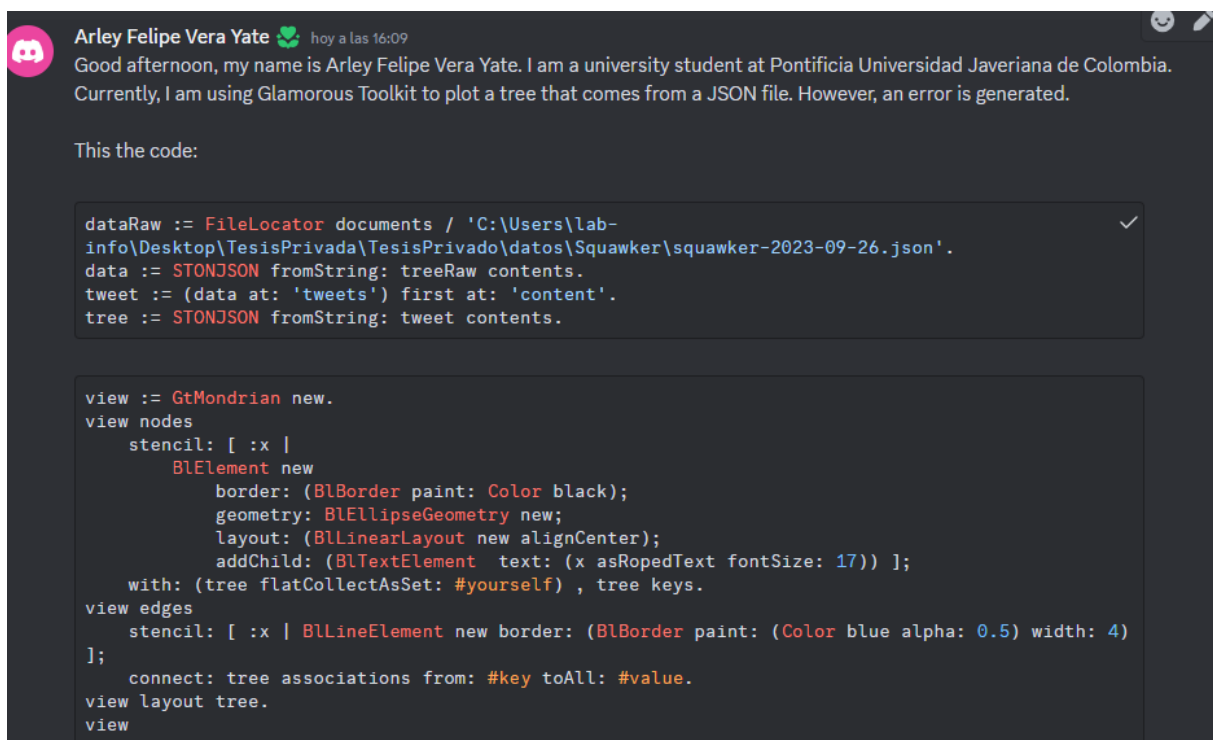


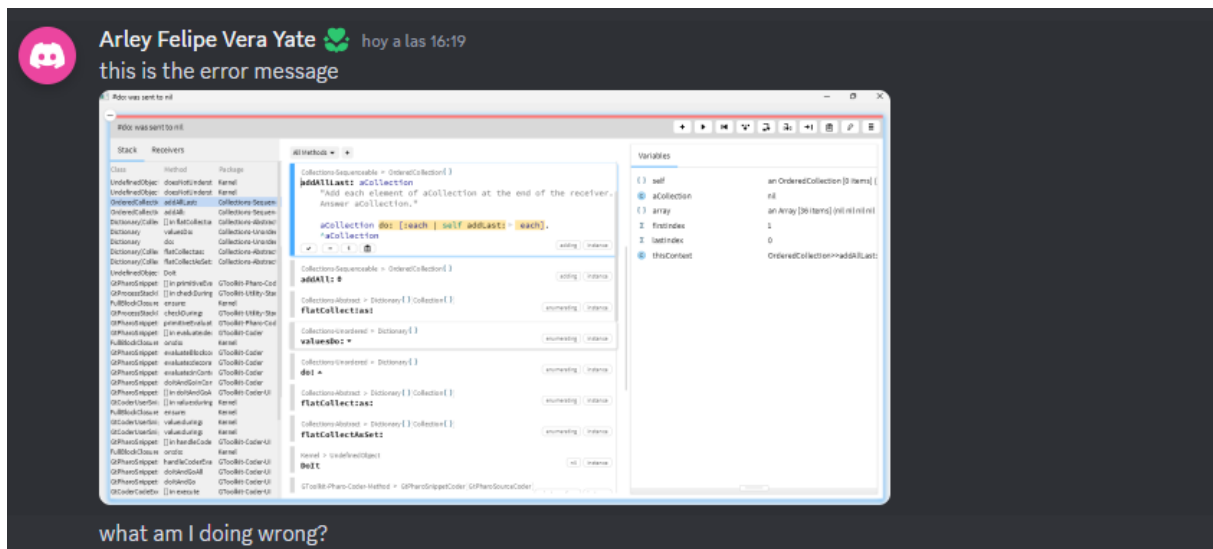
Figure 63: Error al intentar graficar el árbol de datos.

Para el análisis de datos de Squawker, se experimentaron dificultades en la parte de comandos para visualizar los datos que están en formato JSON en forma de árbol. Se estableció comunicación con los desarrolladores de Glamorous Toolkit con el fin de solicitar ayuda para poder graficar los datos.





# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



## Conclusiones y recomendaciones

1. En este estudio, se logró cumplir con el objetivo general de investigar herramientas de scraping para recopilar y analizar datos de Twitter/X, evaluando la calidad de los microdatos de los candidatos. El análisis se desarrolló en dos dimensiones principales: dimensionalidad y densidad, utilizando tres herramientas (Apify, Squawker y Twitter Scraper).

### . Dimensionalidad:

- **Apify:** Capturó una amplia variedad de datos, incluyendo información detallada sobre autores, hashtags y media adjunta en los tweets. Esto permitió obtener una visión completa de los elementos presentes en los tweets y cómo se relacionan entre sí.
- **Squawker:** Proporcionó datos extensos sobre las interacciones de los usuarios, tales como respuestas, retweets y likes. Esta herramienta permitió capturar no solo los datos textuales, sino también el comportamiento de los usuarios en relación con cada tweet.
- **Twitter Scraper:** Recopiló datos textuales de los tweets, hashtags y menciones, ofreciendo una cobertura integral de los contenidos publicados y permitiendo identificar patrones y tendencias en el discurso de los candidatos.

### Densidad:

- **Apify:** Proporcionó una alta densidad de datos sobre autores y hashtags, capturando un volumen considerable de tweets por autor y múltiples hashtags asociados. Además, recogió una cantidad significativa de media (imágenes, videos) adjunta, lo que permitió un análisis detallado de la presencia y el impacto visual en los tweets.
  - **Squawker:** Destacó en la densidad de datos sobre interacciones, recopilando un gran número de respuestas, retweets y likes por cada tweet. Esto permitió un análisis profundo de la participación de los usuarios, mostrando cómo interactuaban con el contenido y qué tipo de tweets generaban más engagement.
  - **Twitter Scraper:** Fue eficiente en obtener una gran cantidad de datos textuales, capturando un volumen significativo de tweets y hashtags, así como menciones y enlaces. Esta densidad de información textualmente rica facilitó la identificación de temas y sentimientos predominantes en el discurso político.
2. En este estudio, se logró cumplir con el primer objetivo específico al configurar un entorno de investigación reproducible para el análisis de datos de Twitter, enfocado en los tweets de los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la Alcaldía de Bogotá. Este entorno reproducible es fundamental por varias razones.

En primer lugar, la recolección y análisis de datos es crucial. Un entorno reproducible permite que todos los pasos realizados durante la investigación sean claros y verificables, garantizando que otros investigadores puedan revisar y validar los resultados obtenidos.

En segundo lugar, la replicabilidad es un aspecto esencial en la investigación. Este entorno reproducible asegura que los métodos y resultados puedan ser replicados por otros investigadores, lo que no solo valida nuestras conclusiones, sino que también facilita estudios comparativos.

Para lograr esto, se integraron diversas herramientas y metodologías que permiten una instalación y configuración estandarizada. Esto incluye la gestión de versiones de los documentos, la automatización de tareas de scraping y análisis.

La implementación de este entorno reproducible no solo ha permitido una recolección y análisis exhaustivos de los microdatos. De este modo, se asegura que tanto la transparencia como la eficiencia y la capacidad de replicación de nuestros métodos y resultados estén garantizadas, cumpliendo así con el primer objetivo específico de la investigación.

3. En el presente estudio, se logró cumplir con el segundo objetivo específico de caracterizar las herramientas de extracción de datos utilizadas (Apify, Squawker y Twitter Scraper) para analizar los tweets de los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante la campaña electoral para la Alcaldía de Bogotá 2023. A continuación, se presentan estas herramientas organizadas de la más a la menos eficiente, destacando también sus defectos.

Squawker se destacó como la herramienta más eficiente y precisa para la captura de datos relacionados con las interacciones de los usuarios, como respuestas, retweets y likes. Esta capacidad permitió una recopilación detallada y diversificada de los datos, proporcionando una rica cantidad de información que facilitó una evaluación de la calidad de los microdatos y ramificarlos. Sin embargo, Squawker puede ser menos intuitivo y más complejo de configurar en comparación con las otras herramientas, y su enfoque en interacciones puede resultar en una menor eficiencia en la recolección de datos textuales y de media.

Apify demostró ser efectiva en la captura de una amplia gama de datos de Twitter, incluyendo texto de los tweets, información de los autores y hashtags. Su capacidad para manejar grandes volúmenes de datos permitió una recolección exhaustiva, y mostró un rendimiento eficiente en la ejecución de tareas de scraping, completando la recolección de datos en un tiempo razonable sin comprometer la calidad. Sin embargo, Apify puede presentar una curva de aprendizaje pronunciada para usuarios nuevos y requiere una configuración inicial detallada.

Twitter Scraper fue altamente eficaz en la recolección de datos textuales y en la identificación de patrones en los tweets. Su enfoque en el contenido textual permitió una cobertura integral de los temas discutidos, operando de manera eficiente y procesando grandes cantidades de

datos textuales en un tiempo reducido. No obstante, esta herramienta tiene limitaciones en la recolección de media adjunta y no captura interacciones como retweets y likes, lo que puede limitar la riqueza y la profundidad de los microdatos recolectados.

4. Ante los cambios abruptos en los términos y condiciones de uso del API en las plataformas de redes sociales, se sugiere, no sólo brindar tiempos extendidos, como aquellos con los que esta tesis contó, gracias a deferencia de la Universidad con estas circunstancias excepcionales, sino también una vinculación temprana de los estudiantes de pregrado a semilleros, talleres y grupos de investigación, que les permita acercarse a problemas investigativos y lidiar de manera ágil y acompañada con las contingencias del acceso a los datos. Por ejemplo, reformulando los problemas de investigación, como se hizo en esta tesis misma. Para ello las metodologías diseñísticas, son muy recomendadas debido a su posibilidad de cambio e iteración permanentes.
  5. Explorar la posibilidad de utilizar nuevas herramientas de scraper para obtener una visión más completa y diversa de la información relacionada con los perfiles de los candidatos en las elecciones. Así mismo, sería beneficioso explorar la posibilidad de utilizar múltiples fuentes de datos además de Twitter para obtener una visión más completa y diversa de la información relacionada con los perfiles de los candidatos en las elecciones
  6. Considerar la posibilidad de ampliar la formación académica en áreas como minería de datos, análisis de redes sociales, metodologías de investigación en Ciencias de la Información y ética en el uso de datos, para abordar de manera más efectiva los desafíos contemporáneos en el análisis de datos y la investigación en redes sociales.
  7. Se recomienda mayor cautela respecto a las capacidades de ChatGPT para la ejecución de instrucciones de análisis de datos y su veracidad en la interpretación de resultados. Esto debido a algunas imprecisiones o falcedades con algunos resultados dados por esta herramienta. Se sugiere no usarla en este ámbito de análisis de datos, al menos que se cuente con experticia de como detectar los fallos de esta herramienta.
- El camino iniciado por esta tesis puede continuarse en al menos dos frentes: cómo usar entornos sencillo de investigación y publicación reproducible y cómo crear narrativas de datos que permitan auditar tanto la calidad de los microdatos extraídos como análisis de discurso e impacto de los candidatos y otras figuras de la vida pública, a partir de sus discurso en redes sociales.

## Bibliografía

- Abdullah, N. H., Hassan, I., Azura Tuan Zaki, T. S., Ahmad, M. F., Hassan, N. A., Mohd Zahari, A. S., Ismail, M. M., & Azmi, N. J. (2022). Examen de la relación entre los factores que influyen en la búsqueda de información política a través de los medios sociales entre los jóvenes de Malasia. *Revista De Comunicación De La SEECI*, 55, 1–15. <https://doi.org/10.15198/seeci.2022.55.e746>.
- Altamirano-Benitez, V., Ruiz-Aguirre, P., & Baquerizo-Neira, G. (2022). Política 2.0 en Ecuador. Análisis del discurso y la comunicación política en Facebook. [Policy 2.0 in Ecuador. Analysis of discourse and political communication on Facebook] *Revista Latina De Comunicación Social*, (80), 201–223. <https://doi-org.ezproxy.javeriana.edu.co/10.4185/RLCS-2022-1539>
- Botero, (2023, 20 octubre). De la plaza pública a las pantallas móviles: así son las campañas electorales de hoy. *Revista Pesquisa Javeriana*. <https://www.javeriana.edu.co/pesquisa/redes-sociales-campanas-electorales/>
- Cela, J. R., Parras-Parras, A., & Romero-Vara, L. (2019). Uso de las redes sociales en diplomacia, política y relaciones internacionales. Análisis de la información publicada en las versiones online de dos periódicos españoles: “el país” y “la vanguardia”. *Estudios Sobre El Mensaje Periodístico*, 25(2), 711–726. <https://doi.org/10.5209/esmp.64798>.
- Croteau, D., & Hoynes, W. (2014). *Media/Society: Industries, Images, and Audiences*. Sage Publications. <https://cadmus.eui.eu/handle/1814/19235>
- D’Agostino, M., Marti, M., Mejía, F., Gerardo, D. C., & Faba, G. (2017). Estrategia para la gobernanza de datos abiertos de salud: un cambio de paradigma en los sistemas de información. <https://iris.paho.org/handle/10665.2/33903>
- Delgadillo, (2022, 21 junio). La campaña política se mueve al ritmo de las redes sociales. *Revista Pesquisa Javeriana*. <https://www.javeriana.edu.co/pesquisa/redes-sociales-campo-disputa-politica/>
- Doria, P., Cortés, S., Doria, P., & Cortés, S. (2023, 19 septiembre). El partidor de candidatos a la Alcaldía de Bogotá. *La Silla Vacía*. <https://www.lasillavacia.com/historias/silla-nacional/el-partidor-de-candidatos-a-la-alcaldia-de-bogota/>
- Front Matter · GitBook. (s. f.). <http://www.practicereproducibleresearch.org/>
- García Acosta, & Gómez Masjuán, M. E. (2022). Fake news en tiempos de posverdad. Análisis de informaciones falsas publicadas en Facebook durante procesos políticos en Brasil y México 2018. Fake news in post-truth times. Analysis of fake news published on Facebook during political processes in Brazil and Mexico 2018 *Estudios Sobre El Mensaje Periodístico*, 28(1), 91–101. <https://doiorg.ezproxy.javeriana.edu.co/10.5209/esmp.71251>

- GmbH, F. (s. f.-a). About. Glamorous Toolkit.<https://gtoolkit.com/about/>
- HOWTO\_chisel: How to chisel. (s. f.).[http://chiselapp.com/user/aplsimple/repository/HOWTO\\_chisel/doc/trunk/](http://chiselapp.com/user/aplsimple/repository/HOWTO_chisel/doc/trunk/)
- Efe, Días, C. (2023, February 2). Twitter empezará a cobrar el acceso de los desarrolladores a su API. Cinco Días.[https://cincodias.elpais.com/cincodias/2023/02/02/companias/1675353625\\_269385.html](https://cincodias.elpais.com/cincodias/2023/02/02/companias/1675353625_269385.html)
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4), 51–58.[https://www.researchgate.net/publication/209409849\\_Framing\\_Toward\\_Clarification\\_o](https://www.researchgate.net/publication/209409849_Framing_Toward_Clarification_o)
- Fernández, A. G. (2018, 17 noviembre). Cambridge Analytica, el big data y su influencia en las elecciones. CELAG.<https://www.celag.org/cambridge-analytica-el-big-data-y-su-influencia-en-las-elecciones/>.
- Grapsas, T. (2021, 12 febrero). Para saciar la curiosidad: ¿conoce la historia de las redes sociales! Rock Content - ES.<https://rockcontent.com/es/blog/historia-de-las-redes-sociales/> .
- Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Science*.<http://www.practicereproducibleresearch.org/>
- Lite XL. (s. f.).<https://lite-xl.com/en/>
- Luna, O. (4 de diciembre de 2023). MetodologiaIntro.<https://uniabi.tiddlyhost.com/#MetodologiaIntro>.
- Marketing Zone Icesi. (2023, 13 marzo). Esta es la historia de Twitter, la app que revolucionó la comunicación en 140 caracteres | Marketing Zone Icesi.<https://www.icesi.edu.co/marketingzone/esta-es-la-historia-de-twitter-la-app-que-revoluciono-la-comunicacion-en-140-caracteres/>^
- McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, 36(2), 176-187. Oxford University Press on behalf of the American Association for Public Opinion Research.<http://www.jstor.org/stable/2747787>
- MutabiT. (s.f.). Socialmetrica.<https://mutabit.com/repos.fossil/mutabit/doc/tip/wiki/es/socialmetrica-28npg.md.html>
- Pérez Curiel, C. y García-Gordillo, M. (2020). Indicadores de influencia de los políticos españoles en Twitter. Un análisis en el marco de las elecciones en Cataluña. *Estudios sobre el mensaje periodístico*, 26 (3), 1133-1144. <https://idus.us.es/handle/11441/102923>.
- Prezzavento, Timoteo. (s.f.). Cambridge Analytica y su Impacto en las Políticas de Privacidad.[https://repositorio.udea.edu.ar/jspui/bitstream/10908/16625/1/\[P\]\[W\]%20T.L.%20Com.%20Prezzavento](https://repositorio.udea.edu.ar/jspui/bitstream/10908/16625/1/[P][W]%20T.L.%20Com.%20Prezzavento)
- Quacker. (2024, 15 mayo). Twitter Scraper · Apify. Apify. <https://apify.com/quacker/twitter-scraper>
- Quijano, H. (2020, 20 agosto). Elementos de las narrativas digitales basadas en datos. Paul Aguilar.<https://socialtic.org/blog/elementos-de-las-narrativas-digitales-basadas-en-datos/>

- Rcn, N. (2023, 9 octubre). Tracking RCN: estos son los candidatos que lideran la intención de voto en Bogotá. Noticias RCN.<https://www.noticiarscn.com/colombia/tracking-rcn-estos-son-los-candidatos-que-lideran-la-intencion-de-voto-en-bogota-455539>.
- ScoopInstaller. (s. f.). GitHub - ScoopInstaller/Scoop: A command-line installer for Windows. GitHub.<https://github.com/ScoopInstaller/Scoop#readme>
- Selva-Ruiz, D., Curbelo, J., & Rodríguez-Méndez, J. (2019). Información de las marcas políticas y su impacto en el rendimiento de la publicidad digital. Information from political brands and its impact on digital advertising performance Estudios Sobre El Mensaje Periodístico, 25(2), 813–830. <https://doi-org.ezproxy.javeriana.edu.co/10.5209/esmp.64799>
- Twitter Scraper. (s. f.).<https://chromewebstore.google.com/detail/twitterscraper/cedomiiokkcmbeoekchahgm>
- Valdebenito Allendes, Jorge. (2018). Twitteo, ¿luego resisto? Movilización popular y redes sociales en Chile: La marea roja de Chiloé (2016). Izquierdas, (40), 185-201. Epub 01 de junio de 2018.<https://dx.doi.org/10.4067/S0718-50492018000300185>.
- Wandmalfarbe. (s. f.). GitHub - Wandmalfarbe/pandoc-latex-template: A pandoc LaTeX template to convert markdown files to PDF or LaTeX. GitHub. <https://github.com/Wandmalfarbe/pandoc-latex-template>
- Zaitsev, O., & Ferlicot-Delbecque, C. (n.d.). Data Analysis Made Simple with Pharo DataFrame.<https://github.com/DataFrame?tab=readme-ov-file>
- Zettlr. (s. f.). *Your One-Stop publication workbench*. Zettlr. <https://www.zettlr.com/>

## Anexos

Objetivos Específicos	Fase según Objetivos	Descripción de la Fase	Actividades para cumplir la Fase	Técnicas para Recoger Información	Resultados o Productos por Fase
1. Configurar un entorno de investigación reproducible para el análisis de datos, centrándose en mensajes (Tweets) relacionados con los candidatos Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la alcaldía de Bogotá.	1. Fase de Configuración del Entorno de Investigación	En esta fase, se establecerá un entorno de investigación reproducible que permita analizar datos específicos de Twitter/X relacionados con Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la alcaldía de Bogotá.	1.1 Seleccionar y configurar herramientas y plataformas para el análisis de datos de Twitter. 1.2 Definir parámetros de búsqueda y filtros para identificar mensajes relacionados con los candidatos. 1.3 Establecer un flujo de trabajo reproducible y documentar la configuración del entorno.	Ninguna	Entorno de investigación configurado y documentado para el análisis reproducible de mensajes relacionados con Gustavo Bolívar y Juan Daniel Oviedo durante las elecciones para la alcaldía de Bogotá.



Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

Objetivos Específicos	Fase según Objetivos	Descripción de la Fase	Actividades para cumplir la Fase	Técnicas para Recoger Información	Resultados o Productos por Fase
2. Caracterizar las herramientas para la extracción de datos “X/Twitter” de Gustavo Bolívar y Juan Daniel Oviedo durante la campaña electoral para la Alcaldía de Bogotá.	2. Fase de Caracterización de Herramientas de Extracción de Datos	En esta fase, se llevará a cabo la caracterización detallada de las herramientas utilizadas para extraer datos de Twitter sobre Gustavo Bolívar y Juan Daniel Oviedo durante su campaña electoral.	2.1 Identificar y describir herramientas específicas utilizadas para la extracción de datos de Twitter. 2.2 Evaluar capacidades y limitaciones en términos de alcance, precisión y velocidad. 2.3 Recopilar información sobre parámetros de búsqueda y filtros utilizados. 2.4 Documentar hallazgos y resultados de la caracterización.	Ninguna	Informe detallado de herramientas de extracción de datos “X/Twitter” utilizadas durante la campaña electoral de Gustavo Bolívar y Juan Daniel Oviedo para la Alcaldía.

El proyecto DataFrame de Pharo proporciona herramientas para la gestión y el análisis de datos estructurados. Admite operaciones de acceso, modificación, enumeración, aritméticas y estadísticas en colecciones de datos. El paquete Pharo DataFrame ofrece métodos para detectar, eliminar y reemplazar valores nulos en una colección de datos, así como leer y escribir en archivos externos. También le permite agregar y agrupar datos para su análisis. Además, el proyecto incluye un tutorial completo de su API, con ejemplos de cada método, así como instrucciones para leer y escribir archivos CSV, que los científicos de datos suelen utilizar para preservar y distribuir datos tabulares.

En este ejemplo, creamos un diccionario llamado dataDensity que simula la densidad de datos para

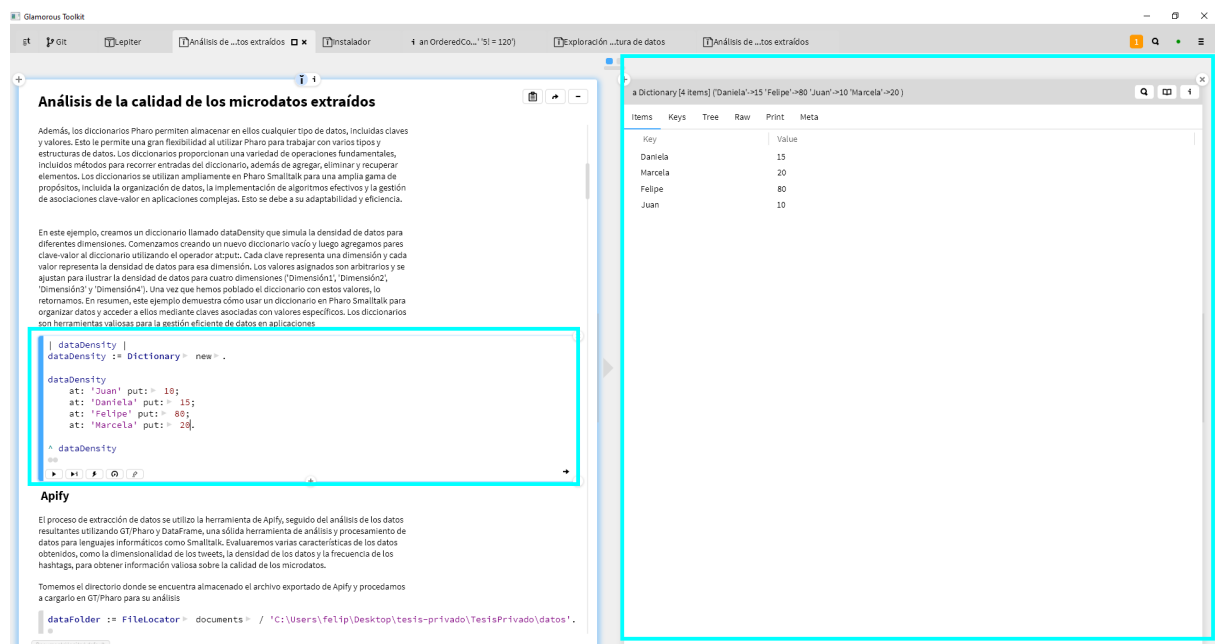
# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

diferentes dimensiones. Comenzamos creando un nuevo diccionario vacío y luego agregamos pares clave-valor al diccionario utilizando el operador `at:put:`. Cada clave representa una dimensión y cada valor representa la densidad de datos para esa dimensión. Los valores asignados son arbitrarios y se ajustan para ilustrar la densidad de datos para cuatro dimensiones ('Dimensión1', 'Dimensión2', 'Dimensión3' y 'Dimensión4'). Una vez que hemos poblado el diccionario con estos valores, lo retornamos. En resumen, este ejemplo demuestra cómo usar un diccionario en Pharo Smalltalk para organizar datos y acceder a ellos mediante claves asociadas con valores específicos. Los diccionarios son herramientas valiosas para la gestión eficiente de datos en aplicaciones

```
| dataDensity |  
dataDensity := Dictionary new.
```

```
dataDensity  
  at: 'Juan' put: 10;  
  at: 'Daniela' put: 15;  
  at: 'Felipe' put: 80;  
  at: 'Marcela' put: 20.
```

```
^ dataDensity
```



## Tesis pregrado Felipe Vera: FileLocator alias tutor

Este archivo permite definir un conjunto de alias rápidos que se cargan en un archivo en localizaciones distintas a partir de un nombre común. De este modo, dos o más personas trabajando en una narrativa

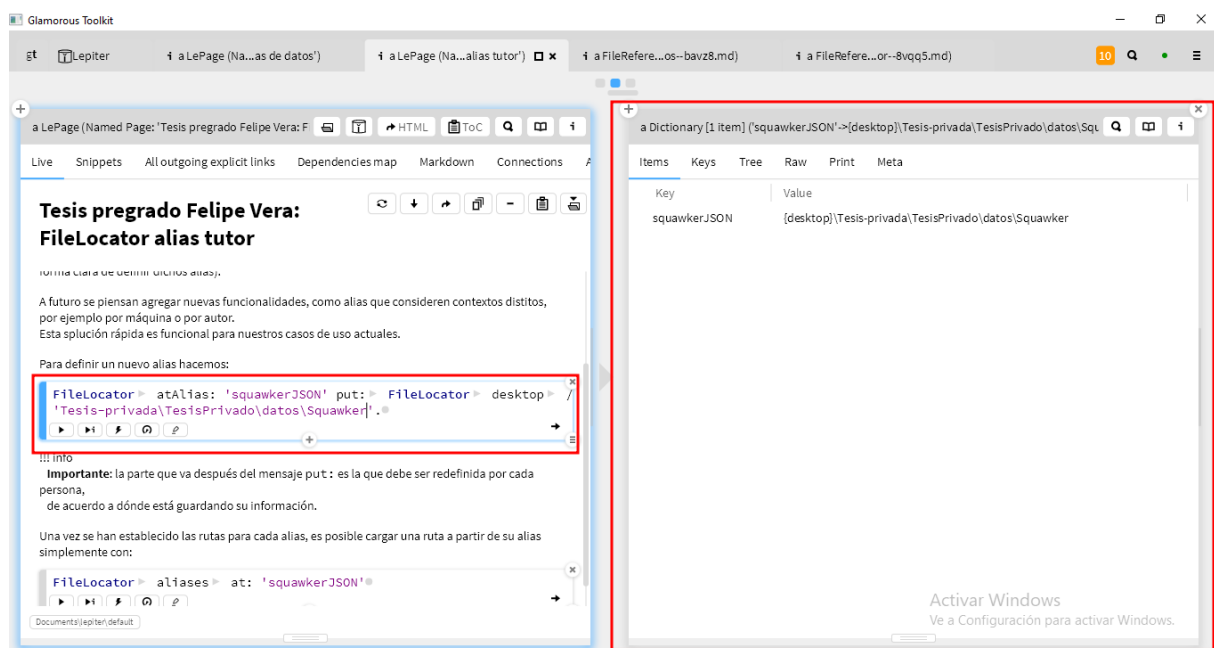
de datos, pueden ejecutarlas sin tener que andar cambiando las rutas particulares a sus archivos, de acuerdo a gustos y organizaciones propias de cada co-autor y tan sólo se solicita que lleguen a un consenso mínimo sobre las alias que le darán a cada ruta.

El archivo que define los alias se guarda en la máquina local de cada coautor y la narrativa que crea dichos alias puede estar o no dentro del control de versiones compartido (aunque se recomienda, cuando lo está, que haya una narrativa distinta por cada coautor o una forma clara de definir dichos alias).

A futuro se piensan agregar nuevas funcionalidades, como alias que consideren contextos distintos, por ejemplo por máquina o por autor. Esta solución rápida es funcional para nuestros casos de uso actuales.

Para definir un nuevo alias hacemos:

```
FileLocator atAlias: 'squawkerJSON' put: FileLocator desktop /  
↔ 'Tesis-privada\TesisPrivado\datos\Squawker'.
```



**Importante:** la parte que va después del mensaje `put:` es la que debe ser redefinida por cada persona, de acuerdo a dónde está guardando su información.

Una vez se han establecido las rutas para cada alias, es posible cargar una ruta a partir de su alias simplemente con:

```
FileLocator aliases at: 'squawkerJSON'
```

## Exploración recursiva de estructuras de datos

En esta narrativa de datos exploraremos cómo implementar la exploración de estructuras de datos de manera recursiva en Pharo. Esto es debido a que los datos en Squawker están guardados en un diccionario anidado o árbol de datos y para poder determinar la diversidad y densidad de los datos, como hemos hecho con otras informaciones tabulares, requerimos primero “aplanar” los datos, lo cual implica recorrerlos de modo recursivo de manera que los hayamos recogido todos en una estructura tabular para aplicar análisis similares a los de otras fuentes de datos.

**Algoritmo del factorial en Pharo** Uno de los problemas matemáticos y de programación más comunes es el cálculo factorial. El producto de todos los números positivos del 1 al  $n$  es el factorial de un entero positivo  $n$ , representado como  $n!$ . El factorial de cinco ( $5!$ ) es, por ejemplo,  $5 \times 4 \times 3 \times 2 \times 1 = 120$ .

**Usos:** El cálculo factorial se utiliza en programación, estadística, matemáticas y técnicas de optimización para abordar problemas de cálculo factorial como permutaciones y combinaciones.

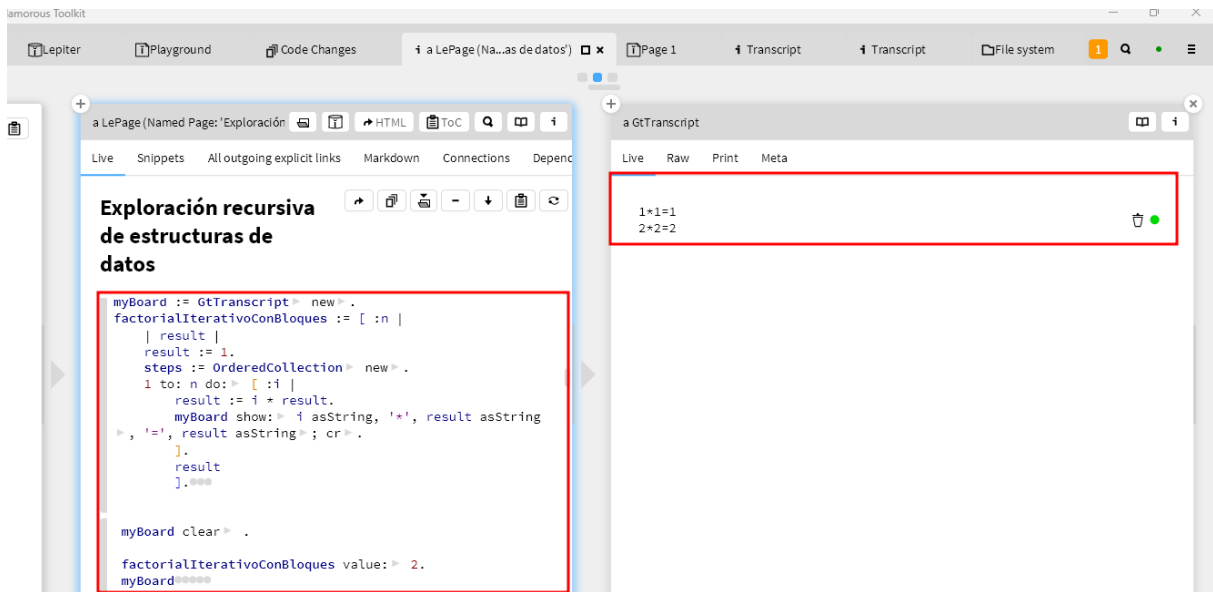
El cálculo factorial resulta útil para determinar de cuántas maneras se puede organizar un elemento o para resolver desafíos combinatorios.

Se utiliza para abordar una variedad de problemas relacionados con el cálculo factorial en programación, estadística, matemáticas y métodos de optimización. Puedes determinar rápidamente el factorial de un número en Pharo sin utilizar la recursividad ejecutando el algoritmo factorial de forma iterativa. Esta implementación es apropiada para circunstancias que requieren un enfoque más sencillo y sencillo porque es sencilla y fácil de entender.

**Implementación iterativa** Una manera eficiente de calcular el factorial de un número en Pharo es mediante un enfoque iterativo. A lo largo de esta narrativa, exploraremos paso a paso cómo implementar este algoritmo.

```
myBoard := GtTranscript new.  
factorialIterativoConBloques := [ :n |  
  | result |  
  result := 1.  
  1 to: n do: [ :i |  
    result := i * result.  
    myBoard show: i asString, '*', result asString, '=', result asString; cr.  
  ].  
  result
```

```
] .  
Smalltalk  
myBoard clear .  
  
factorialIterativoConBloques value: 2.  
myBoard
```



**Figure 64:** FALTA: Agregar leyenda de esta figura

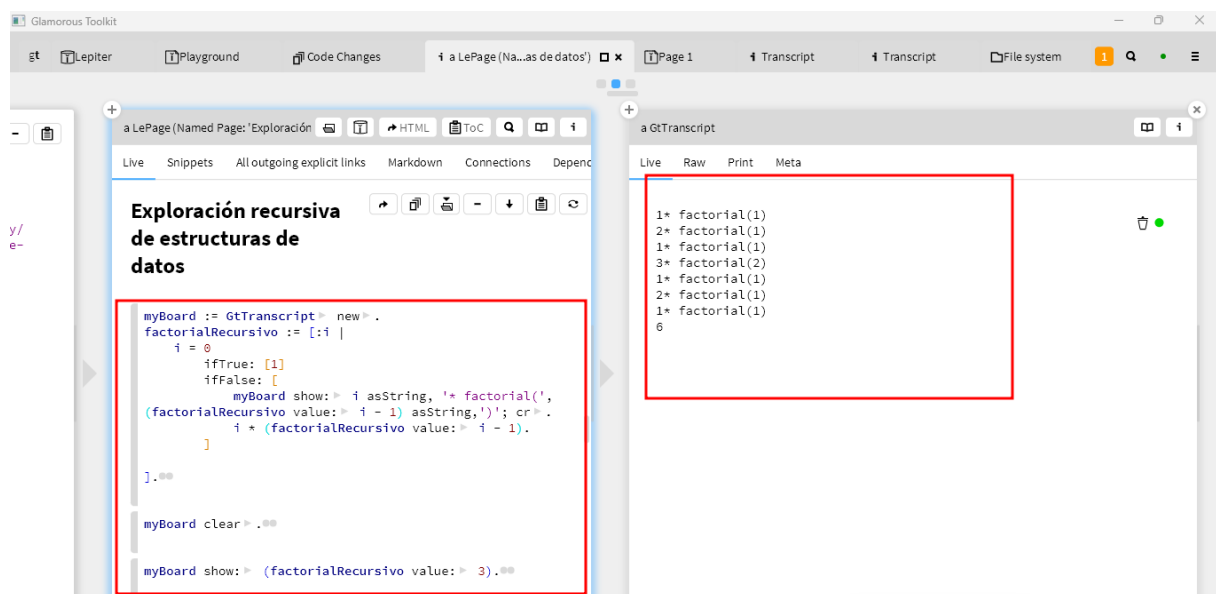
1. “n” se toma como argumento para calcular el numero deseado.
2. Se inicia “result” como 1, porque es el valor inicial al factorial.
3. 1 to: n do: [ :i ]: Este es un bucle que recorre todos los números desde 1 hasta n.
4. result: Finalmente, se devuelve el valor de result, que será e1. “i” se toma como argumento para calcular el numero deseado.

**Implementación Recursiva del Factorial** El objetivo de la implementación factorial recursiva, una técnica de programación popular, es dividir un problema en subproblemas más pequeños y resolver cada uno de ellos de forma recursiva.

Puedes obtener rápidamente el factorial de un número utilizando un método basado en recursividad implementando el algoritmo factorial de forma recursiva en Pharo. Cuando se necesita una respuesta de cálculo factorial simple pero elegante, esta implementación resulta útil.

En la implementación recursiva del factorial, se calcula el factorial de un número dividiendo el problema en subproblemas más pequeños hasta llegar a un caso base.

```
myBoard := GtTranscript new.  
factorialRecursivo := [:i |  
    i = 0  
    ifTrue: [1]  
    ifFalse: [  
        myBoard show: i asString, '* factorial(', (factorialRecursivo value: i -  
1) asString,')'; cr.  
        i * (factorialRecursivo value: i - 1).  
    ]  
].  
  
myBoard clear.  
  
myBoard show: (factorialRecursivo value: 3).
```



1. "i" se toma como argumento para calcular el numero deseado.
2. `i = 0 ifTrue: [1] ifFalse: [...]`: Esta es una expresión condicional. Si `i` es igual a 0, devuelve 1 (porque el factorial de 0 es 1). De lo contrario, ejecuta el código dentro del bloque `ifFalse`.

3. `myBoard show: i asString, '* factorial(', (factorialRecursivo value: i - 1) asString, ')'; cr.:` Esta línea muestra el paso actual del cálculo del factorial en `myBoard`. y el resultado de llamar recursivamente a `factorialRecursivo` con `i - 1`, mostrando este resultado como un string.
4. `i * (factorialRecursivo value: i - 1)`, esto calcula el factorial de `i` multiplicándolo por el factorial de `i - 1`.

**Recorrido de árboles en preorden** En esta sección se reimplementará el recorrido de árboles en preorden, siguiendo el tutorial de Preorder, inorder and postorder tree traversals

Algotree > Algorithms > Tree & Graph Traversal > Preorder, Inorder and Postorder tree traversals

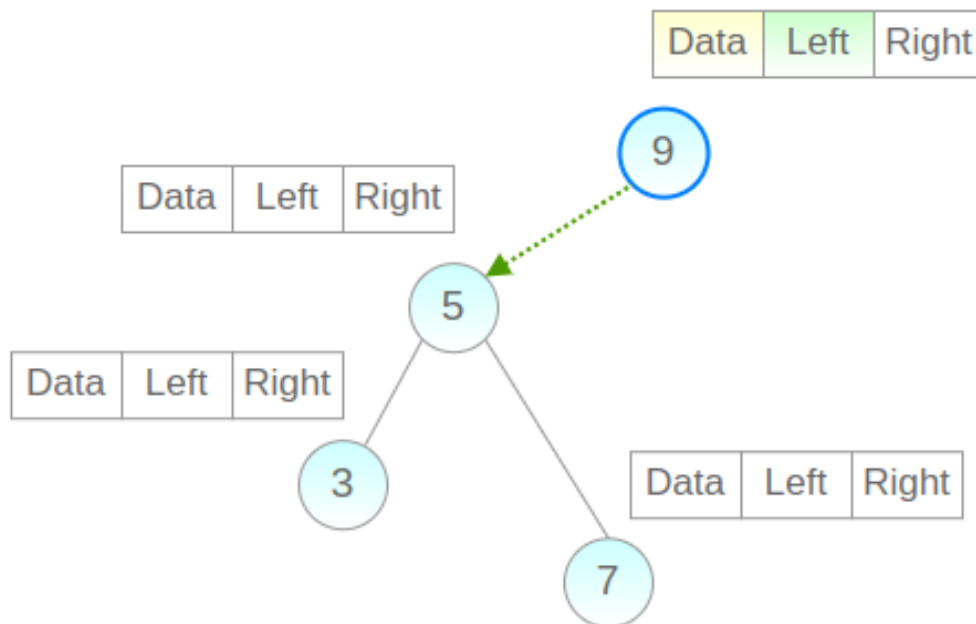
#### PreOrder traversal

- In **Pre-Order** tree traversal, the root data is accessed as soon as the root is visited. After the root data is accessed, the left child of the root node is visited and then the right child.
- The traversal is recursive in nature, i.e. the left child and the right child are traversed similarly to the parent node.
- Thus the preorder traversal recursively follows the sequence: **Print node's data -> Visit Left\_Child\_Of\_Node -> Visit Right\_Child\_Of\_Node.**

#### Algorithm PreOrder Traversal

Recursive	Iterative
<pre>PreOrder ( Node node ) If ( node != null ) then   Print node's data   PreOrder ( node's left child )   PreOrder ( node's right child )</pre>	<pre>PreOrder ( Node node ) Stack stk While ( node != NULL or stk is not empty ) do   If ( node != NULL ) then     Print node's data     Push node on stack stk.     node = node's left child.   If ( node == NULL ) then     node = node at stk top.     Pop stk.     node = node's right child.</pre>

Supongamos que tenemos un árbol como este:

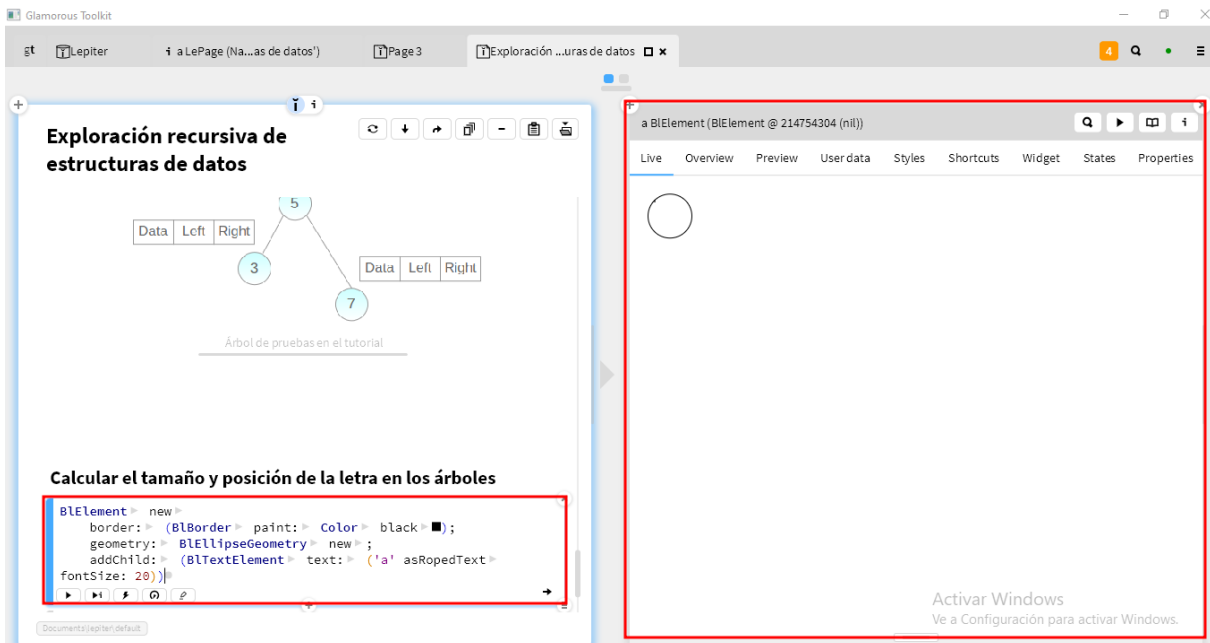


### Calcular el tamaño y posición de la letra en los árboles

```
BElement new
  border: (BtBorder paint: Color black);
  geometry: BtEllipseGeometry new;
  addChild: (BtTextElement text: ('a' asRopedText fontSize: 20))
```

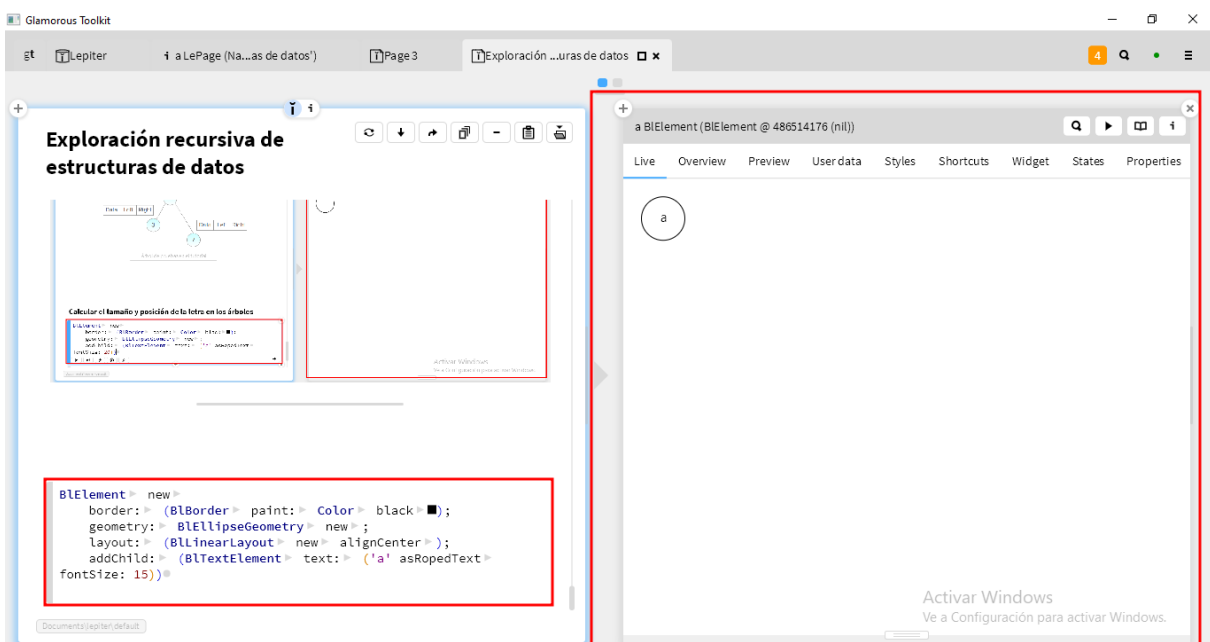


# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



BLElement new

```
border: (BlBorder paint: Color black);  
geometry: BEllipseGeometry new;  
layout: (BlLinearLayout new alignCenter);  
addChild: (BlTextElement text: ('a' asRopedText fontSize: 15))
```

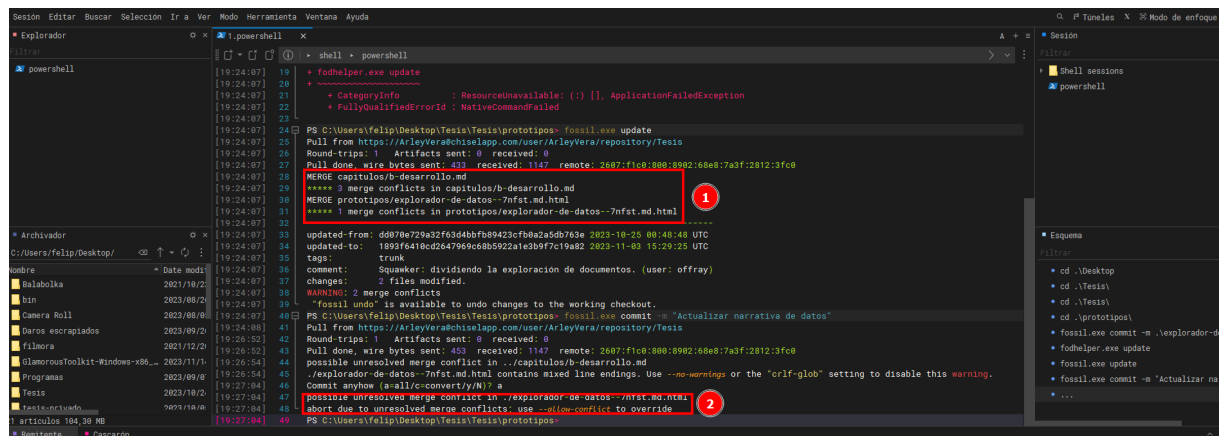


## Uso de chatGPT

Desde su lanzamiento en noviembre de 2022, ChatGPT es una herramienta de inteligencia artificial creada por OpenAI (Telefónica, 2023). Con la ayuda de esta tecnología, las personas pueden interactuar más ampliamente con la IA, lo que permite el soporte por chat para una amplia gama de consultas. ChatGPT tiene múltiples usos, desde escribir textos complejos hasta hacer música o resolver problemas matemáticos. Su desarrollo, desde iteraciones anteriores como Chat GPT-3 hasta el Chat GPT-4 más actual (Telefónica, 2023).

A través del sitio web OpenAI, los usuarios pueden utilizar ChatGPT de forma gratuita, sin embargo, hay una opción premium disponible para más funciones.

Las instrucciones de análisis de datos de Squawker se implementaron de forma más rápida y sencilla gracias a ChatGPT. Pero después de dar los detalles necesarios para completar los pedidos, se encontraron varios problemas. Por ejemplo, se observó que había problemas por la falta de descripciones precisas. Se proporcionaron respuestas, sin embargo, carecían de una justificación.



```
PS C:\Users\Felipe\Desktop\Tesis\Tesis\prototipos> fossil.exe update
Pull from https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
Round-trips: 1 Artifacts sent: 0 received: 0
Pull done, wire bytes sent: 433 received: 1147 remote: 2687:file:888:8902:6808:7a3f:2812:13fc0
MERGE capitulo/b-desarrollo.md
**** 0 merge conflicts in capitulo/b-desarrollo.md
MERGE prototipos/explorador-de-datos-7nfst.md.html
**** 1 merge conflicts in prototipos/explorador-de-datos-7nfst.md.html
updated-from: dd870e729a32f63d4bbf89423cf8a2a5db763a 2023-10-25 00:148:148 UTC
updated-to: 1893f6418cd2647969c68b5922a1e3b9f7c19a82 2023-11-03 15:29:25 UTC
tags: trunk
comment: Squawker: dividiendo la exploración de documentos. (user: ofray)
changes: 2 files modified.
WARNING: 2 merge conflicts
fossil undo is available to undo changes to the working checkout.
PS C:\Users\Felipe\Desktop\Tesis\Tesis\prototipos> fossil.exe commit -m "Actualizar narrativa de datos"
Pull from https://ArleyVera@chiselapp.com/user/ArleyVera/repository/Tesis
Round-trips: 1 Artifacts sent: 0 received: 0
Pull done, wire bytes sent: 433 received: 1147 remote: 2687:file:888:8902:6808:7a3f:2812:13fc0
possible unresolved merge conflict in ../capitulo/b-desarrollo.md
../explorador-de-datos-7nfst.md.html contains mixed line endings. Use --no-warnings or the "crif-glob" setting to disable this warning.
Commit anyhow (s all/c-convert/y/N)? a
possible unresolved merge conflict in ../explorador-de-datos-7nfst.md.html
abort due to unresolved merge conflicts. use --abort-conflict to override
PS C:\Users\Felipe\Desktop\Tesis\Tesis\prototipos>
```

Figure 65: ChatGPT 1

Una parte del código era casi correcta, lo que sugería que otra parte estaba equivocada o no arrojaba el resultado deseado. Este fue un problema adicional que surgió. Esta circunstancia ha resultado en confusión y mayor dificultad para poner en práctica las soluciones sugeridas.

ChatGPT 3.5 ▾

```
smalltalk Copy code

Object subclass: FactorialCalculator [
  FactorialCalculator class >> factorialIterativo: n [
    | resultado i |
    resultado := 1.
    i := 1.
    [i <= n] whileTrue: [
      resultado := resultado * i.
      i := i + 1.
    ].
    ^ resultado
  ]
].

"Uso del método factorialIterativo"
| resultadoFactorial |
resultadoFactorial := FactorialCalculator factorialIterativo: 5.
Transcript show: 'El factorial es: ' , resultadoFactorial printString.
```

**Figure 66:** ChatGPT 2

Adicionalmente se identifico que el uso de herramientas como ChatGPT es útil para acelerar la ejecución de instrucciones de análisis de datos desde su creación en 2022, pero se observaron inconvenientes en descripciones incompletas y respuestas injustificadas, incorrectas o con justificaciones falsas. Aún es necesario que mejore algunos aspectos para una optima eficiencia en este ambito.

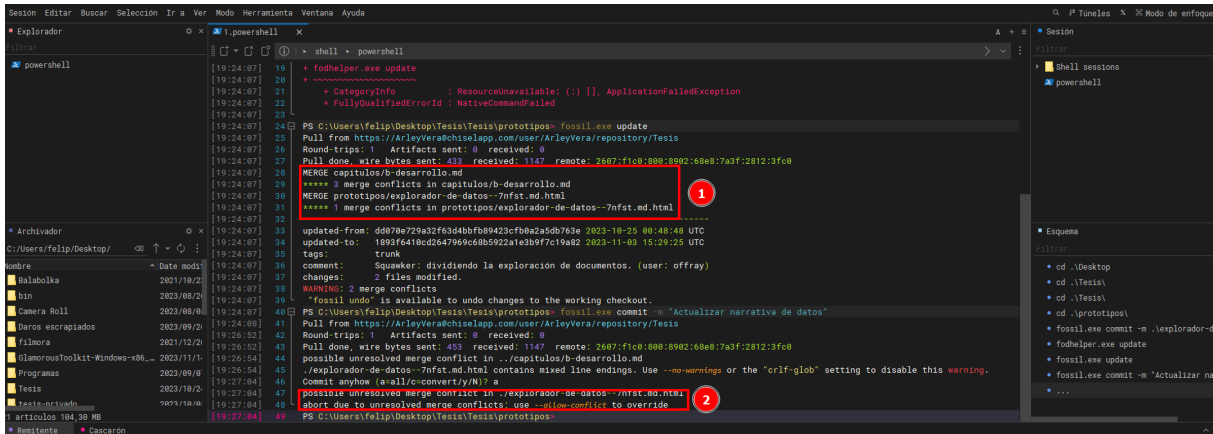
### Diccionarios Pharo Smalltalk

Los diccionarios Pharo Smalltalk son estructuras de datos que le permite emparejar de forma rápida y eficaz claves distintas con valores coincidentes. Los diccionarios Pharo emplean una tabla hash interna para almacenar entradas y están diseñados para ofrecer acceso rápido a la información. Además, los diccionarios Pharo permiten almacenar en ellos cualquier tipo de datos, incluidas claves y valores. Esto le permite una gran flexibilidad al utilizar Pharo para trabajar con varios tipos y estructuras de datos. Los diccionarios proporcionan una variedad de operaciones fundamentales, incluidos métodos para recorrer entradas del diccionario, además de agregar, eliminar y recuperar elementos. Los diccionarios se utilizan ampliamente en Pharo Smalltalk para una amplia gama de propósitos, incluida la organización de datos, la implementación de algoritmos efectivos y la gestión de asociaciones clave-valor en aplicaciones complejas. Esto se debe a su adaptabilidad y eficiencia.

# Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023

## Merge conflicts

En ocasiones, cuando se está trabajando colaborativamente, puede haber conflictos de edición, como se muestra en la siguiente imagen:



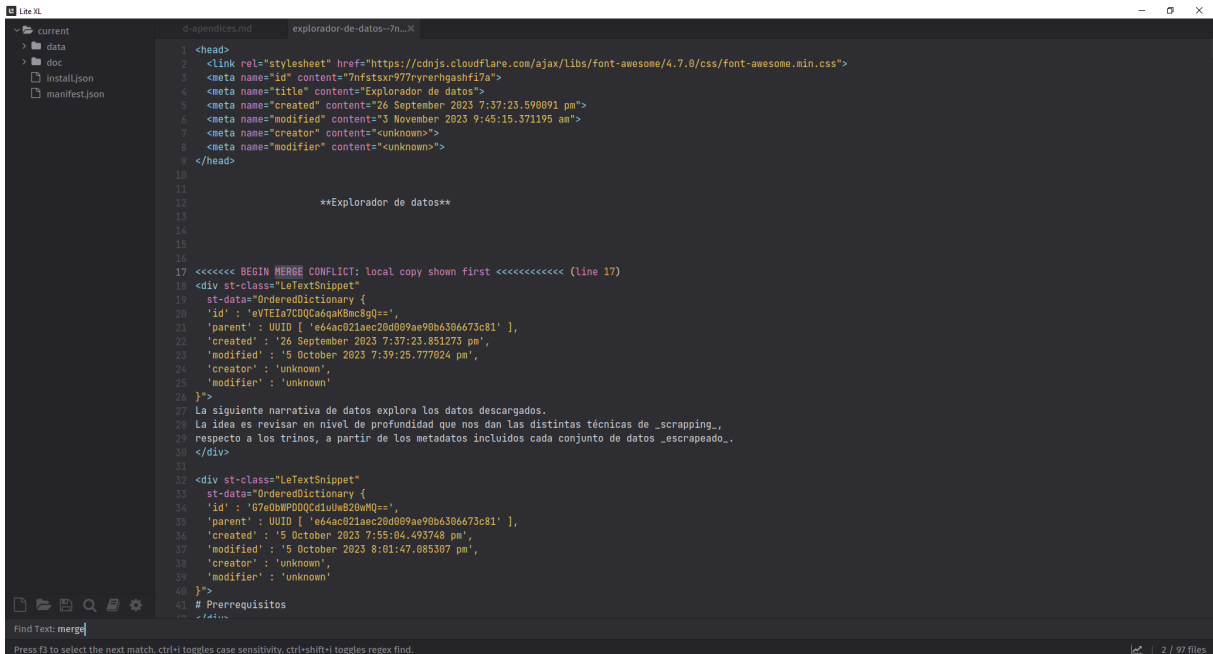
En ocasiones se requiere códigos de texto para ubicar el conflicto de edición y corregirlo manualmente.

Para ello, se instalará el editor “Lite XL”

```
scoop bucket add extras
```

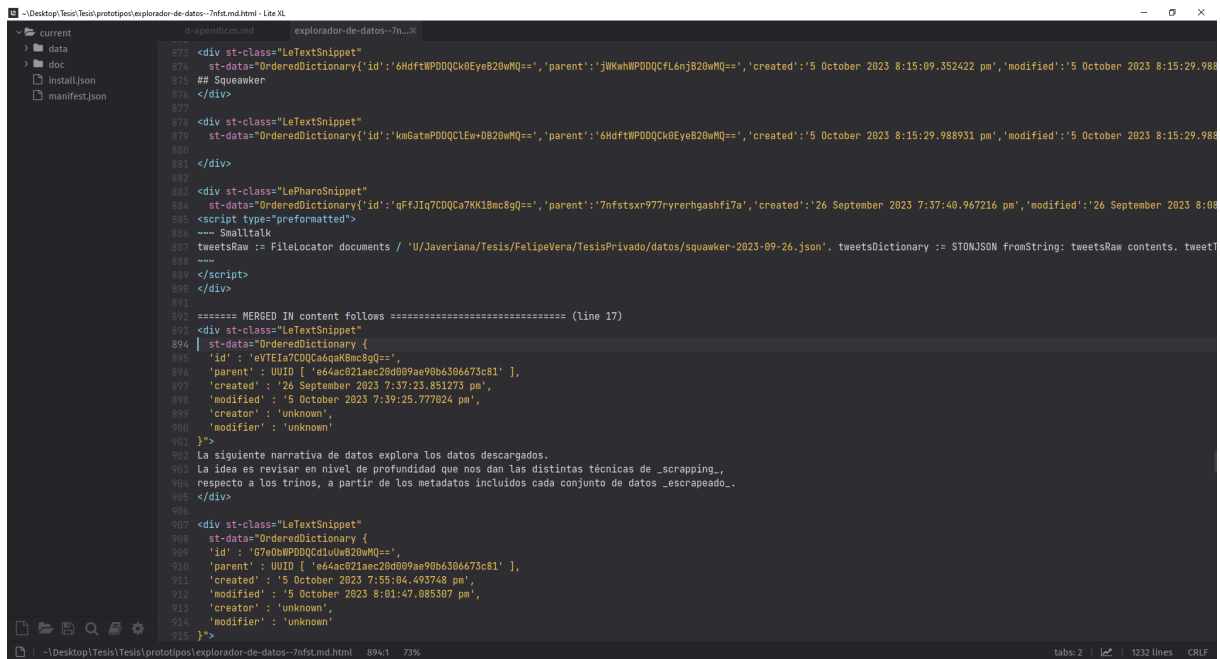
```
scoop install extras/lite-xl
```

Un conflicto de edición se ve así:



Y la propuesta de solución que hace Fossil luce en la siguiente manera

## Análisis de la Calidad de Microdatos Extraídos de Twitter: Un Estudio sobre los Perfiles de los Candidatos Gustavo Bolívar y Juan Daniel Oviedo en las Elecciones para la Alcaldía de Bogotá 2023



```
873 <div st-class="LeTextSnippet"
874   st-data="OrderedDictionary{'id':'6HdfTWPDQCKeY820wMQ==','parent':'jWkWhPPDQCKfL6njB20wMQ==','created':'15 October 2023 8:15:09.352422 pm','modified':'15 October 2023 8:15:29.988
875   # Squeawker
876 </div>
877
878 <div st-class="LeTextSnippet"
879   st-data="OrderedDictionary{'id':'kmGatmPDDQCLEw+0B20wMQ==','parent':'6HdfTWPDQCKeY820wMQ==','created':'15 October 2023 8:15:29.988931 pm','modified':'15 October 2023 8:15:29.988
880 </div>
881
882
883 <div st-class="LePharoSnippet"
884   st-data="OrderedDictionary{'id':'qFfJlq7CDDCa7Kk18mc8gQ==','parent':'7nfstxnr977ryerhghashfi7a','created':'26 September 2023 7:37:40.967216 pm','modified':'26 September 2023 8:08
885 <script type="preformatted">
886 ===== Smalltalk
887 tweetsRaw := FileLocator documents / 'U/Javeriana/Tesis/FelipeVera/TesisPrivado/datos/squeawker-2023-09-26.json'. tweetsDictionary := STONJSON fromString: tweetsRaw contents. tweetT
888 =====
889 </script>
890 </div>
891
892 ===== MERGED IN content follows ===== (line 17)
893 <div st-class="LeTextSnippet"
894   st-data="OrderedDictionary {
895     'id' : 'eVIEIa7CDDCa7Kk18mc8gQ==',
896     'parent' : UUID [ 'e64ac821aec20d089ae98b6306673c81' ],
897     'created' : '26 September 2023 7:37:23.851273 pm',
898     'modified' : '5 October 2023 7:39:25.777024 pm',
899     'creator' : 'unknown',
900     'modifier' : 'unknown'
901   }>
902 La siguiente narrativa de datos explora los datos descargados.
903 La idea es revisar en nivel de profundidad que nos dan las distintas técnicas de _scrapping_,
904 respecto a los trinos, a partir de los metadatos incluidos cada conjunto de datos _escrapeado_.
905 </div>
906
907 <div st-class="LeTextSnippet"
908   st-data="OrderedDictionary {
909     'id' : 'G780WHPDQCKeY820wMQ==',
910     'parent' : UUID [ 'e64ac821aec20d089ae98b6306673c81' ],
911     'created' : '5 October 2023 7:55:04.493748 pm',
912     'modified' : '5 October 2023 8:01:47.085307 pm',
913     'creator' : 'unknown',
914     'modifier' : 'unknown'
915   }>
```

Se puede ver que el conflicto ocurre a raíz de que los metadatos estaban antes en una sola línea y ahora están en varias líneas. Una forma sencilla de arreglarlo, es simplemente sobrescribir el documento con la versión más reciente generada desde Lepiter.