

<https://doi.org/10.15446/mag.v40n1.124493>

**DE LA RECOLECCIÓN MANUAL
DE DATOS A LA AUTOMATIZACIÓN
DIGITAL: UN ENFOQUE HÍBRIDO PARA
LA INVESTIGACIÓN DE FEMICIDIOS**

CARINA ELIZABETH GÓMEZ

Centro de Estudios del Sur Andino (UNJU-CONICET)



*cgomez@fhycs.unju.edu.ar ORCID: [0000-0003-2759-6863](https://orcid.org/0000-0003-2759-6863)

Artículo de investigación recibido: 18 de marzo de 2024. Aprobado: 5 de noviembre de 2025.

Cómo citar este artículo:

Gómez, Carina Elizabeth. 2026. “De la recolección manual de datos a la automatización digital: un enfoque híbrido para la investigación de femicidios”. *Maguaré* 40, 1: 85-120. DOI: <https://doi.org/10.15446/mag.v40n1.124599>

RESUMEN

Este artículo presenta la transición metodológica desde la recolección manual de información hacia la automatización digital mediante *web scraping* en Python, aplicada al estudio de los femicidios en Jujuy, Argentina. La inconsistencia y fragmentación de los registros oficiales y no gubernamentales motivaron el desarrollo de una alternativa metodológica. La recolección manual inicial permitió identificar el problema, aunque con limitaciones significativas en términos de tiempo y capacidad de procesamiento. En contraste, el *scraping* optimizó la recolección de datos periodísticos, ampliando su alcance y sistematización. A partir de ambas experiencias, el artículo propone un enfoque híbrido que, además de superar las restricciones técnicas de los métodos manuales, establece las bases para un protocolo metodológico integral. Este enfoque articula herramientas computacionales con supervisión etnográfica y una perspectiva crítica sobre la violencia feminicida, contribuyendo tanto a su visibilización como a una comprensión más compleja del fenómeno.

Palabras clave: automatización, femicidios, Python, *scraping*.

ABSTRACT

This article discusses the methodological transition from manual data collection to digital automation through web scraping in Python, applied to the study of femicides in Jujuy, Argentina. The inconsistency and fragmentation of both official and non-governmental records prompted the development of an alternative methodological approach. Initial manual data collection allowed for the identification of the problem, albeit with significant limitations in terms of time and processing capacity. In contrast, scraping optimized the collection of journalistic data, expanding both its reach and systematization. Drawing on both experiences, the article proposes a hybrid approach that, in addition to overcoming the technical constraints of manual methods, lays the foundation for an integrated methodological protocol. This approach combines computational tools with ethnographic oversight and a critical perspective on femicidal violence, contributing to both its visibility and a more nuanced understanding of gendered and sexual violence.

Keywords: automation, femicide, Python, *scraping*.

INTRODUCCIÓN¹

La problemática de los femicidios en Argentina, y particularmente en su extremo norte, en la provincia de Jujuy, constituye un fenómeno de gran importancia social que demanda no solo un reconocimiento político y colectivo, sino también el desarrollo de metodologías de investigación que permitan dimensionar y comprender su magnitud.

Es preciso señalar que el femicidio fue definido por Jane Caputi y Diana Russell como el extremo de un *continuum* de violencias que incluye manifestaciones como la violación, la tortura, el abuso sexual infantil, la mutilación genital, el acoso sexual, el maltrato físico y la heterosexualidad forzada (Caputi y Russell 1990, 425). El estudio y documentación de este tipo de violencia extrema enfrenta desafíos significativos, entre los que destaca la disparidad y fragmentación de los registros e informes existentes, lo que dificulta una comprensión integral del fenómeno.

La provincia de Jujuy representa un caso particularmente alarmante en el contexto nacional argentino (Gómez 2020; 2022). Según datos de la Corte Suprema de Justicia de la Nación, esta provincia registró las tasas más altas de femicidio del país en los años 2016 y 2020, ubicándose en segundo lugar en 2017 y en quinto lugar en 2022. Estos indicadores, sumados al mayor índice nacional de violencia doméstica reportado por la Encuesta de Prevalencia de Violencia contra las Mujeres (Ministerio de las Mujeres, Género y Diversidad 2023), configuran un escenario que exige un abordaje metodológico riguroso y multidimensional.

El presente trabajo surge de la constatación de inconsistencias significativas entre los registros de femicidios elaborados por el sector estatal y aquellos producidos por organizaciones de la sociedad civil. Estas discrepancias evidencian problemas teórico-metodológicos, entre ellos la ausencia de un sistema unificado de documentación y análisis que permita dar cuenta de la dimensión real de esta problemática.

¹ El artículo deriva de una parte de los resultados de la tesis doctoral “Entre la vida y la muerte. Las víctimas colaterales del femicidio en la escena pública de Jujuy entre los años 2012-2020” (Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de Jujuy, año 2023) en el marco de una beca doctoral subsidiada por CONICET.

La observación de estas disparidades en un contexto local motivó el desarrollo de una investigación previa que permitió construir una base de datos propia.

En este marco, el artículo tiene como objetivo central describir la transición metodológica desarrollada en dos etapas: 1) una metodología manual de recolección y sistematización de información sobre femicidios (2012-2020) y 2) la implementación de *web scraping* mediante el lenguaje de programación Python para el registro de femicidios (2021-2024). Esto con el fin de diseñar e implementar en el futuro un protocolo metodológico híbrido que combine el *scraping* automatizado con una supervisión etnográfica, permitiendo optimizar los registros frente a los sesgos y fragmentaciones presentes en Jujuy, y ofreciendo un modelo metodológico replicable en otros contextos.

Por “metodología manual” me refiero a la pesquisa de noticias mediante acciones humanas, es decir, la lectura, selección y clasificación de artículos que, aunque cuidadosa, resulta limitada por el tiempo que estas actividades implican. En contraste, la automatización mediante *web scraping*, si bien reduce drásticamente la carga temporal, introduce otros sesgos: los sesgos algorítmicos. Se ha demostrado que los algoritmos pueden perpetuar prejuicios existentes vinculados al género (Benítez 2019). Por lo tanto, esos sesgos derivan de la manera en que los algoritmos priorizan, ordenan y ofrecen la información, reproduciendo prejuicios de género ya presentes en los medios. La transformación metodológica entre ambas etapas responde a la necesidad de optimizar los procesos de recopilación y análisis de grandes volúmenes de datos, superando las limitaciones de los métodos manuales y ampliando las posibilidades analíticas en la investigación social sobre violencias de género. Así, los prejuicios de género presentes en las expresiones mediáticas se traducen en sesgos en los resultados de búsqueda, lo que refuerza la necesidad de un enfoque que articule la labor automatizada con una supervisión etnográfica.

La relevancia de esta propuesta radica en su potencial para contribuir al desarrollo de herramientas metodológicas innovadoras que faciliten la documentación y el análisis sistemático de femicidios, aportando no solo al campo académico de los estudios de género, sino también a la generación de información confiable para la prevención y erradicación de la violencia femicida.

Acerca del recuento de femicidios en Argentina

En 2008, Amnistía Internacional elaboró un informe sobre femicidios ocurridos en Argentina, basado en casos relevados en medios de comunicación. Concluyó que durante ese año se produjeron al menos 100 muertes de mujeres en el país a manos de un miembro de su entorno familiar o de una pareja (Amnistía Internacional 2008). La metodología implementada para este informe se fundamentó en la observación sistemática de casos de violencia publicados en la prensa y en medios digitales durante el período comprendido entre enero y diciembre de ese año. Los medios examinados incluyeron publicaciones de alcance nacional como los diarios *Clarín*, *La Nación*, *Página 12* y *Crónica*, junto con periódicos de relevancia regional como *La Gaceta* (Tucumán), *Los Andes* (Mendoza), *La Voz del Interior* (Córdoba), *El Diario de la República* (San Luis), *El Periódico Austral* (Santa Cruz), *Nuevo Diario* (Santiago del Estero) y *Diario Norte* (Chaco).

Ese mismo año, la Asociación Civil La Casa del Encuentro comenzó el conteo de víctimas de femicidio en el país a través de su Observatorio Adriana Marisel Zambrano (OAMZ), mediante informes estadísticos basados en una metodología similar a la adoptada por Amnistía Internacional (Rico y Tuñez 2013, 9).

También en 2008, el diario *Clarín* desarrolló un relevamiento paralelo que identificó 171 mujeres asesinadas. Aunque no detalló exhaustivamente las fuentes consultadas, el artículo indicaba que el relevamiento se realizó a partir de publicaciones en medios digitales (*Redacción Clarín*, 25 de noviembre de 2008).

En este mismo escenario, es importante señalar los trabajos previos realizados por la periodista Gabriela Barcaglione, quien contabilizó víctimas como “una actividad rutinaria y en algún sentido individual y casi privada” (Barcaglione 2005, 78). En su artículo reflexionó –años antes incluso de la publicación de un primer registro de femicidios– sobre la urgencia de una práctica no solo colectiva, sino también institucional de contabilización de víctimas.

Durante este período, Argentina recibió observaciones de diversos organismos internacionales que señalaron deficiencias en materia de género, entre ellas la carencia de datos oficiales sobre la magnitud de las formas de violencia en general y del femicidio en particular (Defensoría General de la Nación 2015). Sin embargo, a pesar de dichas

observaciones, el país solo desarrolló un registro oficial de femicidios en 2015, año que coincide con un momento de intensificación en la movilización social en torno a la problemática.

El 3 de junio de 2015 marcó un hito con la realización de un acto multitudinario en la Plaza del Congreso de la Nación bajo la consigna “Ni Una Menos”, movimiento que nació, según sus organizadoras, “ante el hartazgo por la violencia machista, que tiene su punto más cruel en el femicidio” (Ni Una Menos, 3 de junio de 2015). En su Manifiesto n.º 1, proclamado en esa misma fecha, el colectivo expresaba:

En 2008 mataron una mujer cada 40 horas; en 2014, cada 30. En esos 7 años, los medios publicaron noticias sobre 1.808 femicidios. ¿Cuántas mujeres murieron asesinadas solo por ser mujeres en 2015? No lo sabemos (...) No existe en la Argentina un registro oficial de femicidios. Los datos que mencionamos aquí son los que releva año tras año la Organización No Gubernamental Casa del Encuentro. Es fundamental contar con estadísticas oficiales (...) Decimos Ni Una Menos, con el dolor de la suma de víctimas que crece y crece. Esta conmoción masiva, esta enorme y comprometida participación social, son un grito unánime. (Ni Una Menos, 3 de junio de 2015)

Uno de los principales reclamos expresados en este manifiesto fue precisamente la necesidad de recopilar y publicar estadísticas oficiales sobre la violencia hacia las mujeres, incluyendo específicamente registros e índices de femicidios. La movilización no solo evidenció estas carencias institucionales, sino que convocó tanto a la sociedad civil como a los organismos gubernamentales a implementar medidas urgentes para abordar la problemática.

Así, el período comprendido entre el primer informe de femicidios de Amnistía Internacional (2008) y la primera marcha Ni Una Menos (2015) estuvo marcado por la transformación jurisprudencial y por la paulatina, pero consistente, visibilización social de la violencia de género. Esta creciente conciencia se desarrolló en conjunto con el esfuerzo sostenido de organizaciones de la sociedad civil, dependencias institucionales especializadas y ámbitos académicos –principalmente del sector periodístico– comprometidos con la temática.

La contabilización sistemática de víctimas de femicidio contribuyó decisivamente al reconocimiento inicial de la magnitud de la violencia

extrema ejercida contra las mujeres; paralelamente, la movilización social facilitó la identificación de la necesidad imperativa de contar con diagnósticos cuantitativos precisos. Estos esfuerzos permitieron comenzar a cubrir significativos vacíos de información y delinear los trazos generales del fenómeno en el contexto nacional argentino.

Finalmente, unos meses después de la histórica marcha Ni Una Menos, el Poder Judicial publicó los primeros datos oficiales sobre femicidios en el país. Así, luego de tres décadas de haber ratificado la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer, el Estado argentino desarrolló bancos de datos, investigaciones y estadísticas que permitieron conocer oficialmente la magnitud de la problemática.

Sin embargo, para 2015, la Corte Suprema de Justicia de la Nación (CSJN) informó un total de 235 víctimas (Corte Suprema de Justicia de la Nación 2016), mientras que el OAMZ registró un número superior: 286 femicidios y femicidios vinculados de mujeres y niñas, así como 43 femicidios vinculados de hombres y niños (La Casa del Encuentro, s.f.). Esta divergencia evidencia una problemática que abordo a continuación.

Como fenómeno social, el femicidio se encuentra atravesado por múltiples actores y relaciones complejas. Esta complejidad se traduce en diversas formas de conceptualización y comprensión del fenómeno y, por lo tanto, en una multiplicidad de abordajes, entre los cuales la aproximación estadística ocupa un lugar privilegiado, particularmente por constituir un insumo central en la formulación de políticas públicas.

Como expuse hasta aquí, desde 2008 Argentina ha sido escenario de diversos esfuerzos institucionales y sociales dirigidos a la identificación del femicidio como problemática social y a la contabilización de víctimas. Esta iniciativa inicial fue posteriormente complementada por el surgimiento de registros alternativos desarrollados por organizaciones de la sociedad civil y colectivos de activistas feministas. Un ejemplo de esta labor es el Observatorio de la organización Mujeres de la Matria Latinoamericana (MuMaLa), que desde 2015 mantiene un registro sistemático de femicidios y otras muertes violentas de mujeres y disidencias. En este panorama también destacan, aunque no son objeto de análisis exhaustivo en este trabajo, los registros elaborados por el Observatorio Ahora Que Sí Nos Ven y el Observatorio Lucía Pérez.

Esta pluralidad de aproximaciones metodológicas ha derivado en una marcada disparidad en los resultados de los estudios estadísticos sobre femicidios en Argentina. En su primer informe, la propia CSJN confirmó estas dificultades:

(...) pese al esfuerzo en aras de compatibilizar los datos pertenecientes a diversos sistemas de registro para la elaboración de estas estadísticas, la pluricausalidad y complejidad de la problemática que implica la intervención de múltiples instituciones, diversidad de enfoques y criterios de registro, así como la heterogeneidad en el grado de sistematización de la información y cuestiones de competencia inherentes a la organización federal de nuestro país, no han permitido aún la unificación de la información.

A ello, se ha sumado una dificultad adicional producto de la posterior incorporación de nuevos requerimientos específicos del MESECVI que contiene más de 50 indicadores estructurales de proceso y de resultado que deben ser informados para dar cuenta del cumplimiento de la Convención de Belém do Pará y en la que debe unificarse información perteneciente al ámbito nacional, federal y provincial de la Justicia. (Corte Suprema de Justicia de la Nación, 2014, 5)

De lo anterior se confirma que, en cada informe, se manifiestan modalidades heterogéneas de contabilización, interpretación e intervención sobre los femicidios. Las diferencias en las cifras entre los distintos registros responden a múltiples factores. En primer lugar, existe una ausencia de definiciones comunes: los registros analizados emplean tipologías diversas. Mientras algunos distinguen entre femicidios directos y vinculados –por ejemplo, el OAMZ diferencia *Femicidios Vinculados de mujeres y niñas* y *Femicidios Vinculados de hombres y niños* (Rico y Túñez 2013, 25)–, la CSJN distingue entre *Femicidios Vinculados* y *Femicidios Vinculados por Interposición en la línea de fuego* (Corte Suprema de Justicia de la Nación, s.f.).

Respecto de los asesinatos de mujeres trans y travestis, las categorías también varían: el OAMZ considera *Femicidios a Mujeres Trans o Travesticidios* (Asociación Civil La Casa del Encuentro, 2020, 65), mientras que la Defensoría del Pueblo de la Nación (DPN) utiliza la categoría *Femicidios Trans* (Defensoría General de la Nación, 2018). Asimismo, algunos registros incorporan categorías adicionales como

lesbicidios (Mujeres de la Matria Latinoamericana, 2020) o *suicidios feminicidas* (Defensoría General de la Nación, 2018). Esta diversidad implica que un mismo caso puede clasificarse de manera distinta según el registro o, directamente, no ser incluido. Sin una categoría de registro compartida, los totales resultan inevitablemente divergentes.

A lo anterior se suma la diversidad de fuentes y criterios de inclusión utilizados por los distintos observatorios. Algunos priorizan información judicial con sentencias firmes; otros combinan datos institucionales y prensa; y los observatorios de la sociedad civil realizan rastreos en medios y fuentes alternativas.

Del mismo modo, la heterogeneidad en las variables registradas plantea otro obstáculo significativo. No todos los registros documentan las mismas dimensiones del fenómeno: algunos reportan edad, nacionalidad u ocupación de las víctimas. Por ejemplo, todos indican la edad, aunque solo la CSJN informa la nacionalidad. La situación procesal, el vínculo con el agresor o la desagregación geográfica tampoco son variables universales, lo que dificulta cualquier comparación más allá del conteo agregado.

A su vez, los cambios metodológicos a lo largo del tiempo introducen dificultades adicionales. Algunos observatorios modifican su cobertura, dejan de publicar datos provinciales o ajustan sus criterios en respuesta a nuevas exigencias –como las del Mecanismo de Seguimiento de la Convención de Belém do Pará–. Estas transformaciones fracturan las series temporales: un año puede registrarse bajo ciertos criterios y el siguiente bajo otros, generando inconsistencias aparentes. Por otra parte, las perspectivas políticas y epistemológicas divergentes afectan tanto la selección de casos como la priorización de categorías. Esta dimensión trasciende lo técnico: aquello que se considera relevante registrar responde también a motivaciones políticas y a formas particulares de comprender la violencia de género.

Las limitaciones prácticas en el acceso a la información constituyen otro factor determinante. El acceso desigual a distintas fuentes, junto con la alta carga de trabajo que implica el rastreo manual, genera subregistros diferenciales. Con el tiempo, los observatorios con acceso a fuentes oficiales, mayores recursos o mejor infraestructura logran series más completas que aquellos con restricciones técnicas.

Finalmente, la dependencia problemática de la prensa como fuente merece atención. Aunque no es el objetivo central de este estudio, es preciso señalar que muchas organizaciones civiles construyeron “contra-datos” (D’Ignazio et ál., 2020) o “datos discordantes” (Suárez Val, 2020) a partir de medios de comunicación ante la ausencia de registros estatales o ante los subregistros existentes. Es claro que esta vía tiene límites: los archivos periodísticos no siempre son completos, en muchos casos revictimizan a las mujeres víctimas de femicidio, carecen de perspectiva de género u omiten información. Recurrir a la prensa como fuente no garantiza automáticamente mayor precisión, pero aporta valor si se demuestra que amplía la cobertura o permite triangular información, tal como lo haría una metodología híbrida, objeto de este artículo.

PRIMERA ETAPA: RECOLECCIÓN MANUAL DE DATOS

Contexto y fundamentación teórico-metodológica

En la fase inicial de esta investigación desarrollé un estudio comparativo de tres observatorios de femicidio de alcance nacional, particularmente relevantes en el contexto argentino: a) Observatorio Adriana Marisel Zambrano, b) Corte Suprema de Justicia de la Nación Argentina y c) Observatorio Mujeres, Disidencias y Derechos (MuMaLa). Esta etapa preliminar de búsqueda de documentación cuantitativa sobre femicidios se caracterizó por constituir un campo fragmentario, disperso en diversos informes. Esta heterogeneidad se manifiesta no solo en la multiplicidad de fuentes emisoras, sino también en la diversidad metodológica empleada para la obtención de datos y en los distintos formatos de publicación adoptados.

El conjunto analítico final comprendió más de 30 informes anuales correspondientes al período 2012-2020, abordados mediante una metodología que integra dos niveles complementarios de análisis. El primer nivel consistió en un examen cuantitativo centrado en las variaciones numéricas de víctimas reportadas por cada observatorio. El segundo nivel profundizó en un análisis cualitativo de los procedimientos metodológicos implementados para la construcción de cada registro, identificando tanto convergencias como divergencias en las estrategias de recolección y sistematización de datos.

Esta aproximación a los datos reveló, desde un inicio, problemas estructurales significativos: discrepancias en las cifras reportadas, fragmentación temporal en los registros y ausencia de criterios unificados para la identificación de casos. Estas limitaciones evidencian un problema metodológico fundamental en la documentación de femicidios en Argentina que obstaculiza la comparabilidad de los datos.

En la siguiente tabla comparo los datos de femicidios reportados por cada observatorio entre 2012 y 2020. A primera vista, se observa que la cantidad de víctimas informadas no coincide en ningún registro en ningún año.

Tabla 1. Datos comparados sobre femicidios en Argentina (2012-2020).

Fuente Año	Observatorio Adriana M. Zambrano	Corte Suprema de Justicia de la Nación	Defensoría del Pueblo	Observatorio MuMaLa
2012	255	-	-	-
2013	295	-	-	-
2014	277	225	-	-
2015	286	235	-	232
2016	290	254	-	322/279*
2017	295	251	292	304
2018	273	255	281	298/260*
2019	299	252	280	287
2020	300	251	295	270

Fuente: elaboración propia.

En el caso de las cifras marcadas con asterisco (322 víctimas para el año 2016 y 298 víctimas para el año 2018), se trata de datos oficiales publicados en informes anuales de registro sistemático. Sin embargo, el propio organismo MuMaLa, en un informe posterior titulado “*7 años del primer #NiUnaMenos*”, reportó otros valores para esos mismos años: 279 víctimas en 2016 y 260 víctimas en 2018.

Este patrón trasciende las fronteras nacionales, replicándose sistemáticamente en diversas regiones donde convergen problemas como

la insuficiencia de estudios estadísticos rigurosos, la circulación de narrativas que contribuyen a la invisibilización del fenómeno, deficiencias estructurales en el procesamiento de denuncias y la ausencia de protocolos estandarizados para la investigación y el tratamiento judicial de estos casos (Almérás y Calderón 2012; Gambetta 2022).

En su primer informe la propia Corte Suprema de Justicia confirmó estas dificultades:

pese al esfuerzo en aras de compatibilizar los datos pertenecientes a diversos sistemas de registro para la elaboración de estas estadísticas, la pluricausalidad y complejidad de la problemática que implica la intervención de múltiples instituciones, diversidad de enfoques y criterios de registro, así como la heterogeneidad en el grado de sistematización de la información y cuestiones de competencia inherentes a la organización federal de nuestro país, no han permitido aún la unificación de la información.

A ello, se ha sumado una dificultad adicional producto de la posterior incorporación de nuevos requerimientos específicos del MESECVI que contiene más de 50 indicadores estructurales de proceso y de resultado que deben ser informados para dar cuenta del cumplimiento de la Convención de Belém do Pará y en la que debe unificarse información perteneciente al ámbito nacional, federal y provincial de la Justicia (Corte Suprema de Justicia de la Nación, 2014, 5.)

Estas claras limitaciones evidencian un problema metodológico fundamental en la documentación de femicidios en Argentina, que obstaculiza la comparabilidad de los datos y, por lo tanto, motivó el desarrollo de una metodología alternativa para construir un corpus más completo y consistente, esta vez en un contexto regional.

Contexto y problematización de los registros existentes

Mi investigación sobre femicidios en la provincia de Jujuy se inició mediante un análisis exhaustivo de registros tanto estatales como aquellos elaborados por organizaciones de la sociedad civil. Al examinar los datos proporcionados por los principales sistemas de documentación para el período 2012-2020, identifiqué discrepancias significativas en la información disponible. Los registros del Observatorio Adriana M. Zambrano, la Corte Suprema de Justicia de la Nación, la Defensoría

del Pueblo y el Observatorio MuMaLa produjeron cifras divergentes entre sí, caracterizadas por la ausencia de tendencias estadísticas claras. Esta inconsistencia se ve agravada por la presencia de vacíos informativos para determinados años, lo que compromete la posibilidad de establecer un panorama preciso sobre la evolución del fenómeno en el territorio jujeño.

Como ya desarrollé, un aspecto particularmente problemático de estos registros es su fragmentación temporal, especialmente al momento de pesquisar los datos provinciales. Mientras el Observatorio Adriana M. Zambrano fue el único que proporcionó información para 2012 y 2013, dejó de ofrecer registros a escala provincial a partir de 2017. Por su parte, la Corte Suprema de Justicia comenzó a registrar sistemáticamente los femicidios recién desde 2014, y el Observatorio MuMaLa no proporcionó datos específicos para la provincia de Jujuy en 2020, limitándose a indicar una tasa de 0,8 femicidios sin detallar el número absoluto de víctimas.

La Tabla 2 registra los datos de femicidios que cada entidad informa para la provincia de Jujuy entre 2012 y 2020. En este caso, las problemáticas ya identificadas en los registros nacionales se repiten a nivel provincial; sin embargo, estos números presentan algunas coincidencias anuales.

Tabla 2. Datos comparados sobre femicidios en la provincia de Jujuy (2012-2020).

Fuente Año	Observatorio Adriana M. Zambrano	Corte Suprema de Justicia de la Nación	Defensoría del Pueblo	Observatorio MuMaLa
2012	6	-	-	-
2013	8	-	-	-
2014	3	3	-	-
2015	6	4	-	4
2016	9	10	-	10
2017	-	11	10	8
2018	-	5	5	3

2019	-	3	4	3
2020	-	11	11	-
Total	32	47	30	14

Fuente: elaboración propia.

Las cifras correspondientes a 2017 a 2020 para el Observatorio Adriana M. Zambrano no están indicadas porque, como ya expliqué, desde esa fecha, la entidad no ofrece registros a escala provincial. Otra aclaración sustancial es que el *Registro nacional de Femicidios, Femicidios Vinculados, Trans/Travesticidios y Lesbicidios* de MuMaLa de 2020 determinó que la tasa de femicidios para la provincia de Jujuy fue del 0,8. Este cálculo se realizó “en base a los casos de femicidios directos de mujeres + femicidios vinculados mujer + trans/travesticidios en base a la población estimada al 2020 por el INDEC según proyecciones elaboradas en base al Censo Nacional de Población y Viviendas 2010” (Mujeres de la Matria Latinoamericana 2020, 2). Sin embargo, no indica la cantidad de víctimas registradas para ningún territorio.

Las inconsistencias detectadas revelan un problema metodológico estructural en la documentación de femicidios en Argentina, el cual se manifiesta igualmente a nivel provincial. Como ya advertí, este problema se caracteriza por la ausencia de criterios unificados para la identificación y registro de casos, la discontinuidad en los procesos de recopilación informativa y la implementación de metodologías heterogéneas que obstaculizan significativamente la comparabilidad de los datos obtenidos.

Ante este escenario de fragmentación informativa y evidente subregistro, surgió la necesidad de desarrollar una metodología alternativa –objeto central del presente artículo– orientada a la construcción de un corpus de datos más completo y consistente, que permitiera superar las limitaciones inherentes a los registros existentes y ofrecer una aproximación lo más completa posible al fenómeno estudiado.

Recolección manual de datos

Ante las limitaciones identificadas en los registros disponibles, opté por desarrollar una metodología manual de recopilación y sistematización de información. Esta se fundamentó en la búsqueda y el análisis exhaustivo de noticias publicadas en medios digitales de comunicación locales,

regionales y nacionales. Esta estrategia siguió un proceso sistemático que incluyó diversas etapas, que desarrollo a continuación.

En primer lugar, procedí a la selección de las fuentes periodísticas. Para ello, identifiqué los principales medios digitales que cubren noticias de la provincia de Jujuy, incluyendo medios locales (El Tribuno, Jujuy al Momento, Jujuy al Día, Somos Jujuy, El Submarino, Todo Jujuy, Jujuy es Noticia, El Intransigente, Jujuy Registrado, Jujuy On Line y Qué Pasa Jujuy), medios de provincias cercanas (El Tribuno de Salta, El Intransigente de Salta y La Gaceta de Tucumán) y medios nacionales (Página 12, Clarín, La Nación, Infobae, Cosecha Roja y Revista Anfibia).

En segundo lugar, implementé un diseño de estrategias de búsqueda mediante el motor de Google. Por un lado, apliqué configuraciones de filtros temporales delimitados al período 2012-2020, lo que permitió circunscribir los resultados al marco cronológico específico de la investigación. Paralelamente, utilicé operadores de búsqueda avanzada que posibilitaron una mayor precisión en los resultados obtenidos. Estos operadores incluyeron el uso de comillas para buscar frases exactas, operadores booleanos (AND, OR, NOT) para establecer relaciones lógicas entre términos, comodines para contemplar variaciones terminológicas y operadores de dominio para delimitar la búsqueda a medios específicos. Asimismo, incorporé términos relacionados con la localización geográfica (Jujuy) combinados con descriptores vinculados a la problemática estudiada, particularmente: “femicidio”, “femicida”, “violencia de género”, “asesinato de mujer”, “violencia contra la mujer”, “muerte de mujer”, “crimen de mujer” y “violencia de género”. Esta metodología permitió optimizar tanto la cantidad como la calidad de los resultados, privilegiando aquellas fuentes con mayor relevancia para los objetivos de la investigación.

En tercer lugar, revisé cada resultado relevante individualmente, determinando su pertinencia para la investigación y procediendo a su documentación o eliminación.

En un cuarto procedimiento, clasifiqué la información. Sometí las noticias recopiladas a un análisis de contenido para extraer datos relevantes sobre cada caso de femicidio, incluyendo información sobre la víctima, el victimario, las circunstancias del hecho y el tratamiento mediático del caso.

Estructuré la información analizada en una base de datos que permitió la organización sistemática y el posterior análisis de los datos. Para este

proceso, identifiqué y procesé los siguientes campos: a) nombre de la víctima, b) edad de la víctima, c) fecha del femicidio, d) localidad donde sucedió el crimen, e) nombre del o los agresores, f) edad o edades del o los agresores, g) relación de la víctima con el o los agresores y h) breve descripción del caso.

Finalmente, en quinto lugar, realicé una validación cruzada. Es decir, contrasté la información de distintas fuentes sobre un mismo caso para verificar la solidez de los datos recopilados.

Resultados de la Etapa 1

Este proceso, que implicó la documentación y el análisis manual de más de 550 artículos periodísticos publicados entre 2012 y 2020, requirió un esfuerzo de sistematización significativo, traducido en años de trabajo. La magnitud de la tarea evidenció las limitaciones inherentes de un procedimiento manual, particularmente en términos de tiempo, recursos humanos y posibilidades de procesamiento de grandes volúmenes de información.

En números absolutos, el resultado de esta primera metodología de búsqueda de información fue:

Tabla 3. Datos comparados sobre femicidios en Argentina de acuerdo con la creación de una base de datos propia (2012-2020).

Año	Número
2012	3
2013	5
2014	2
2015	5
2016	11
2017	9
2018	4
2019	3
2020	10
Total	52

Fuente: elaboración propia.

Esta metodología permitió identificar 52 víctimas de femicidio en la provincia de Jujuy durante el período estudiado, una cifra superior a la reportada por cualquiera de los registros estatales o de organizaciones de la sociedad civil. Esta información, desagregada por año, muestra una distribución que no coincide plenamente con la tendencia sugerida por los registros existentes, lo que refuerza la hipótesis de subregistro y fragmentación informativa.

Limitaciones y desafíos metodológicos

A pesar de los aportes significativos de esta primera etapa, la recolección manual de datos tuvo diversas limitaciones que motivaron la búsqueda de metodologías complementarias. Entre los principales desafíos identificados encontré:

a) Alta demanda de tiempo y recursos humanos:

El proceso manual de búsqueda, lectura, análisis y documentación de noticias requirió una inversión considerable de tiempo. Esto dificultó la posibilidad de ampliar la investigación hacia períodos más extensos o hacia otras provincias.

b) Dificultades para procesar grandes volúmenes de información:

La capacidad individual para revisar y sistematizar noticias impuso un límite al volumen de información recuperable y analizables. Esto también incrementó la necesidad de almacenamiento virtual para conservar los archivos recopilados.

c) Sesgos asociados a los motores de búsqueda:

El uso de Google introduce sesgos propios de sus algoritmos de rastreo, indexación y personalización. Como explica Benítez Eyzaguirre (2019), el motor ordena más de 60 trillones de páginas mediante algoritmos que emplean alrededor de doscientas mil variables, además de parámetros de personalización basados en localización, historial de búsqueda, idioma y tipo de navegador. Desde 2013, estos sistemas incorporan criterios como relevancia social, calidad del contenido y preferencias individuales. No obstante, tales mecanismos pueden reproducir prejuicios de género y raciales.

d) Limitaciones en la recuperación histórica de noticias:

Cuanto más antiguo era el período de búsqueda, mayor fue la dificultad para encontrar noticias, especialmente en medios locales

de menor audiencia cuyo archivo digital es limitado o inexistente. Esto afectó particularmente los años 2012 a 2014.

e) Dificultad para la actualización continua del corpus:

El procedimiento manual no resultó eficiente para mantener la base de datos al día con nuevos casos, lo que afectó su sostenibilidad en el tiempo.

Estas limitaciones llevaron a la necesidad de explorar técnicas metodológicas que permitieran agilizar, ampliar y optimizar la recopilación de datos, sin perder rigor en el análisis.

SEGUNDA ETAPA: IMPLEMENTACIÓN DE WEB SCRAPING MEDIANTE LENGUAJE DE PROGRAMACIÓN PYTHON

Fundamentación teórico-metodológica

La transición desde una metodología manual hacia un enfoque basado en *web scraping* responde directamente a las limitaciones identificadas en la etapa anterior. Este cambio metodológico se inscribe en las discusiones contemporáneas sobre los métodos digitales en la investigación social y acerca de las posibilidades que ofrecen las herramientas computacionales para abordar fenómenos complejos documentados en entornos digitales.

El *web scraping*, entendido como el proceso automatizado de extracción de información estructurada desde sitios web mediante software o código de programación (Vilkova 2020), posee un enorme potencial para la investigación social. En estudios sobre femicidios –donde la información se encuentra dispersa en múltiples portales de noticias, con distintos criterios de publicación y archivo– esta técnica permite construir bases de datos más completas, actualizadas y comparables. Aun así, como advierten Luscombe, Dick y Walby (2022), recopilar exhaustivamente contenido de un sitio no garantiza necesariamente representatividad del fenómeno estudiado, lo que exige un análisis crítico posterior.

La implementación de esta técnica no reemplaza la perspectiva desarrollada en la primera etapa, sino que la complementa. La propuesta consiste en articular ambos enfoques para aprovechar:

- la precisión cualitativa del análisis manual, y
- la eficiencia y escalabilidad del procesamiento automatizado.

De este modo, el *web scraping* deja de entenderse únicamente como una herramienta técnica para convertirse en una pieza central de un enfoque metodológico integral. Esta integración se inscribe en un campo emergente que explora el uso de herramientas de programación para analizar fenómenos sociales que se manifiestan, circulan o son documentados en entornos digitales. En este marco, la automatización libera tiempo para profundizar en la lectura crítica, la interpretación contextual y el análisis cualitativo de los casos registrados.

Diseño e implementación del scraping

El diseño experimental inicial se orientó específicamente a la extracción sistemática de noticias sobre femicidios correspondientes al período 2021-2024 del periódico digital *El Tribuno de Jujuy*, uno de los principales medios de comunicación de la provincia y con un archivo histórico amplio y accesible en línea. La elección de este medio como caso piloto respondió a criterios de relevancia regional, accesibilidad del contenido histórico y características estructurales del sitio web que facilitaban la aplicación de técnicas de *web scraping*.

El proceso de diseño e implementación del scraper lo organicé en diversas fases. En primer lugar, realicé un análisis preliminar exhaustivo del sitio web, identificando los patrones de organización de la información, la arquitectura de URLs y la estructura del Document Object Model (DOM) de las páginas de noticias, con especial atención a la sección *Policiales*. Este análisis, apoyado por herramientas de Inteligencia Artificial, fue fundamental para comprender la lógica de presentación digital de los contenidos y, a partir de ello, diseñar estrategias de extracción viables y eficientes.

En una segunda fase, definí con precisión los elementos informativos a extraer de cada noticia:

- a) título,
- b) fecha de publicación,
- c) contenido textual,
- d) autoría,
- e) categoría,
- f) etiquetas asociadas, y

g) elementos multimedia pertinentes.

La delimitación clara de estos objetivos permitió orientar el desarrollo técnico hacia las necesidades analíticas específicas de la investigación.

Para la implementación técnica seleccioné el lenguaje Python, dada su versatilidad, accesibilidad y disponibilidad de librerías especializadas para *web scraping*. Las principales bibliotecas empleadas fueron:

- Requests, para realizar peticiones HTTP;
- BeautifulSoup, para el análisis del HTML y la extracción estructurada de datos;
- csv, para la escritura de los datos obtenidos;
- Datetime, para el manejo de fechas;
- Logging, con el fin de registrar eventos durante la ejecución;
- Time, para introducir pausas controladas entre peticiones y reducir la carga sobre los servidores; y
- Urllib.parse, para la manipulación y normalización de URLs.
- La solución adoptó una arquitectura modular, con componentes especializados para:
 - la navegación sistemática por el sitio,
 - la identificación de enlaces pertinentes,
 - la extracción estructurada de la información de cada artículo, y la exportación de los datos a formatos adecuados para su análisis posterior.

Esta modularidad facilitó el desarrollo incremental del *scraper* y estableció las bases para su futura adaptación a otros medios.

Un eje central del proyecto fue el desarrollo de mecanismos de filtrado que permitieran identificar noticias vinculadas específicamente con femicidios. Para ello diseñé algoritmos que combinaron búsqueda por palabras clave, análisis de frecuencia de términos y clasificación temática, lo que permitió priorizar contenidos relevantes dentro del conjunto más amplio de noticias policiales.

Paralelamente, desarrollé una estructura optimizada de base de datos para almacenar la información extraída de manera ordenada, robusta y accesible. El diseño contempló tanto la integridad de los datos como su disponibilidad para distintos tipos de análisis,

e incorporó metadatos que documentaban el proceso de extracción y sus parámetros técnicos.

Finalmente, un componente sustantivo del desarrollo fue la implementación de prácticas éticas de scraping (Jayachandran y Arni 2023). Incorporé medidas destinadas a garantizar el respeto por los términos de servicio del sitio web de *El Tribuno*, evitando generar cargas indebidas sobre sus servidores. Estas medidas incluyeron:

- la introducción sistemática de pausas entre peticiones,
- la limitación de la frecuencia de acceso, y
- un sistema de registro que permitiera monitorear el comportamiento del *scraper* y ajustar dinámicamente su ejecución para minimizar su impacto.

Resultados de la etapa 2

Estructuré el código resultante en torno a varias funciones principales que operan de manera secuencial, desde la identificación inicial de URLs relevantes hasta el almacenamiento final de la información procesada. La bibliografía muestra que investigadoras e investigadores que extraen información de acceso público suelen compartir su código como complemento en línea de sus informes de investigación, permitiendo que futuras pesquisas lo reutilicen para recopilar la misma información o realicen pequeñas modificaciones para obtener datos similares de las mismas plataformas (Li, Zhou y Cai 2019; Luscombe, Dick y Walby 2022). Por ello, el código desarrollado está disponible públicamente en GitHub, en consonancia con el potencial del rastreo web para fomentar la apertura, el intercambio y su aplicación en otros contextos.

Como se observa en el diagrama de flujo (Figura 1), el proceso inicia con la configuración del sistema de registro (*logging*), que permite realizar un seguimiento detallado de la ejecución y almacenar dicha información. Este registro facilita la identificación de errores y el monitoreo del progreso general del programa.

El flujo principal del *scraper* comienza estableciendo una URL base para la sección de noticias policiales del periódico y configurando encabezados HTTP que simulan un navegador web, con el fin de evitar que el sitio lo bloquee. El programa utiliza un sistema de paginación que inicia en la primera página y avanza sucesivamente hasta que no

encuentre más artículos o alcance contenido correspondiente a años anteriores al buscado.

Para cada página, el *scraper* descarga el contenido HTML y emplea la librería BeautifulSoup para analizarlo y extraer todos los elementos de tipo *article*. Si no se detectan más artículos, el programa finaliza la búsqueda. Para cada artículo encontrado, se extrae la URL y se analiza la fecha contenida en ella para verificar si corresponde al año especificado (2024 por defecto). Esta verificación temporal es fundamental, dado que el objetivo del ejemplo es recopilar únicamente noticias de ese año en particular.

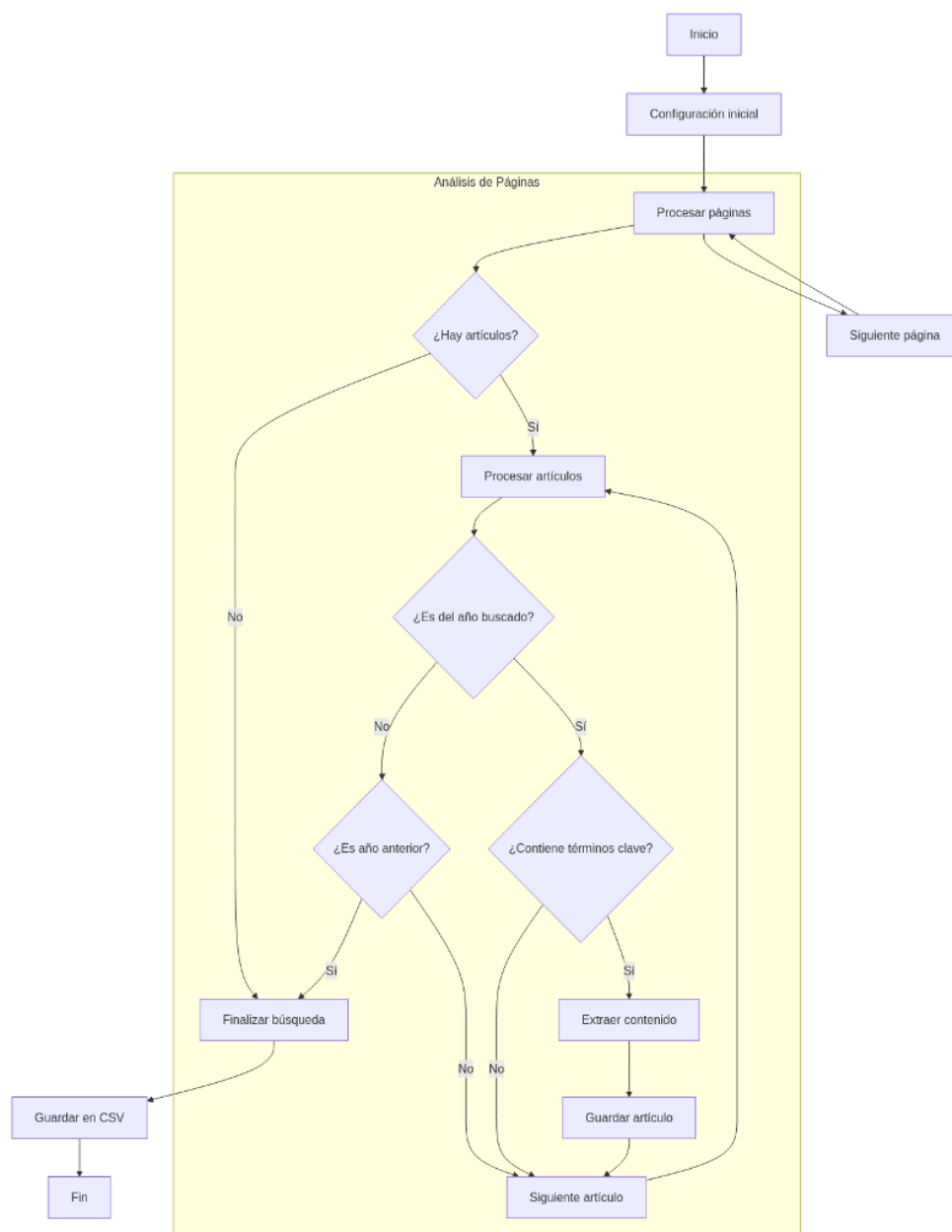
Cuando un artículo pertenece al año de interés, el programa extrae su título y comprueba si contiene alguno de los términos predefinidos asociados a femicidios, tales como “violencia de género”, “asesinato de mujer”, “femicidio”, entre otros. Esta lista es amplia para captar las diversas formas en que los medios pueden referirse a estos casos, especialmente considerando el posible sesgo en la cobertura periodística.

Si el título contiene alguno de estos términos clave, el *scraper* accede a la página individual del artículo para extraer su contenido completo. Durante este proceso se implementan pausas controladas (2 segundos entre artículos individuales y 3 segundos entre páginas) con el propósito de evitar sobrecargar el servidor del medio y disminuir la probabilidad de bloqueo.

Los datos extraídos de cada artículo considerado relevante –título, enlace, contenido, URL y fecha– se almacenan automáticamente. Finalmente, una vez procesadas todas las páginas pertinentes, el programa guarda la información recopilada en un archivo csv nombrado según el año analizado (por ejemplo, *tribuno_femicidios_2024.csv*).

A lo largo de todo el proceso, el *scraper* implementó un sistema de manejo de excepciones destinado a responder a errores potenciales, tales como problemas de conexión, modificaciones en la estructura del sitio web o cualquier otro imprevisto. Esto garantizó que el programa pudiera continuar funcionando o, en caso necesario, finalizar de manera controlada.

Figura 1. Diagrama de flujo del scraper.



Fuente: elaboración propia.

Continuando con el ejemplo, en el siguiente registro del *log* se observa la hora de inicio del *scraper*, correspondiente a las 19:09:25 hs. (Figura 2). A continuación, presento las líneas de código donde se registró el primer hallazgo de noticias de femicidios para el año 2024, cuya marca temporal corresponde a las 19:11:02 hs., es decir, apenas dos minutos después del inicio del proceso (Figura 3).

Figura 2. Registro del log al iniciar el scraping.

```
2025-01-31 19:09:25,813 - INFO - Scraping página 1: https://eltribunodejujuy.com/seccion/policial
2025-01-31 19:09:28,228 - INFO - No se encontraron artículos del año 2024 en la página 1
```

Fuente: elaboración propia.

Figura 3. Registro del log del primer hallazgo de artículo de noticia.

```
2025-01-31 19:10:59,913 - INFO - Scraping página 18: https://eltribunodejujuy.com/seccion/policiales/18
2025-01-31 19:11:02,304 - INFO - No se encontraron artículos del año 2024 en la página 18
2025-01-31 19:11:05,304 - INFO - Scraping página 19: https://eltribunodejujuy.com/seccion/policiales/19
2025-01-31 19:11:07,196 - INFO - Scrapeando artículo: https://eltribunodejujuy.com/edicion-impresa/2024-12-31-0
```

Fuente: elaboración propia.

Luego de ejecutar el *scraper* entre la página 1 y la 19 (donde ocurre la primera extracción de contenido), corroboré manualmente que cada página de la sección *Policiales* de El Tribuno de Jujuy contiene 12 noticias; es decir, el *scraper* analizó al menos 228 noticias en apenas dos minutos. Se trata de un nivel de automatización imposible de alcanzar mediante una metodología manual.

Ventajas y potencialidades del enfoque de programación

La implementación de técnicas de *web scraping* mediante lenguaje de programación Python para la recopilación sistemática de noticias sobre femicidios ofrece ventajas significativas respecto de la recolección manual de datos previamente empleada. Entre sus principales potencialidades destaca la eficiencia en la recopilación: mediante la escritura de líneas de código y el uso de herramientas de código abierto (libre) es posible crear un proceso automatizado que reduce drásticamente el tiempo necesario para obtener información, permitiendo procesar

grandes volúmenes de noticias en lapsos considerablemente menores que los métodos manuales.

Asimismo, la automatización facilita sustancialmente la ampliación del alcance investigativo. Mientras la aproximación manual limitaba de manera inevitable el número de fuentes y el período temporal analizable, el enfoque de programación permite extender sistemáticamente el análisis a múltiples medios periodísticos, intervalos cronológicos más amplios y diversas áreas geográficas. Esta expansión multidimensional incrementa significativamente la representatividad de los datos recopilados y, en consecuencia, fortalece la solidez de los registros que puedan derivarse.

En el mismo ejemplo, presento a continuación las primeras y últimas líneas del *log* (archivo de registro) del *scraper*:

Figura 4. Registro del log al iniciar el scraping.

```
2025-01-31 19:09:25,813 - INFO - Scraping página 1: https://eltribunodejujuy.com/seccion/policiales
2025-01-31 19:09:28,228 - INFO - No se encontraron artículos del año 2024 en la página 1
```

Fuente: elaboración propia.

Figura 5. Registro del log al finalizar el scraping.

```
2025-01-31 19:47:31,962 - INFO - Scraping página 267: https://eltribunodejujuy.com/seccion/policiales/267
2025-01-31 19:47:34,289 - INFO - Scrapeando artículo: https://eltribunodejujuy.com/nota/2024-1-1-17-2-0-br
2025-01-31 19:47:35,692 - INFO - Artículo guardado: Brutal crimen en la costa: una patota emboscó y mató a
2025-01-31 19:47:37,694 - INFO - Scrapeando artículo: https://eltribunodejujuy.com/nota/2024-1-1-13-48-0-cl
2025-01-31 19:47:40,030 - INFO - Artículo guardado: Chascomús: hallan muerto al hombre buscado por el doble
2025-01-31 19:47:42,035 - INFO - Se alcanzó un año anterior (2023), finalizando búsqueda
2025-01-31 19:47:42,078 - INFO - Se encontraron 155 artículos relacionados con femicidios en 2024
2025-01-31 19:47:42,079 - INFO - Los resultados se guardaron en tribuno\_femicidios\_2024.csv
```

Fuente: elaboración propia.

Así como en el ejemplo anterior, es posible observar que entre las páginas 1 y 267 de la sección *Policiales* se realizaron aproximadamente 3.204 extracciones de noticias, de las cuales se identificaron 155 artículos vinculados a los *tokens* preestablecidos. Este proceso de automatización tuvo una duración aproximada de 38 minutos.

Un beneficio adicional es la sistematización y estructuración inherente al tipo de información extraída mediante *web scraping*. A diferencia de la recopilación manual, en la que la organización de los

datos constituye habitualmente un proceso posterior, el *scraper* permite obtener información directamente estructurada según parámetros predefinidos. Esta estructuración nativa facilita de manera sustancial el procesamiento y análisis posterior de los datos.

El enfoque de programación también contribuye a reducir ciertos sesgos específicos asociados a la recopilación manual. Por un lado, disminuye el impacto del agotamiento cognitivo que puede surgir del tratamiento de grandes volúmenes de información; por otro, mitiga parcialmente la influencia algorítmica de los motores de búsqueda sobre los resultados visibles, una limitación difícil de evitar en los procedimientos manuales.

No obstante, resulta fundamental reconocer que la implementación de *web scraping* introduce simultáneamente otros desafíos relacionados con la configuración de los criterios de extracción, cuya definición algorítmica requiere una reflexión crítica. En este sentido, como criterio de rigurosidad, los *tokens* incluidos en el script fueron las mismas palabras clave utilizadas en la primera etapa de búsqueda manual: “femicidio”, “femicida”, “violencia de género”, “asesinato de mujer”, “violencia contra la mujer” y “muerte de mujer”. A estas se sumaron nuevas palabras clave: “violencia machista”, “violencia familiar”, “abuso sexual”, “asesinato”, “crimen”, “mató”, “asesinó”, “hallaron muerta”, “encontraron muerta”, “cadáver de mujer” y “cuerpo de mujer”.

Otra ventaja operativa relevante es la posibilidad de realizar procesos de actualización continua. El *scraper* puede programarse para ejecutarse periódicamente según intervalos predefinidos, lo que permite mantener la base de datos continuamente actualizada con nuevos casos reportados por los medios. Esta capacidad de monitoreo sistemático resulta especialmente valiosa para fenómenos sociales como el campo de estudios sobre femicidios, en el cual la temporalidad y la actualización permanente de la información son claves para comprender patrones emergentes y evaluar intervenciones.

Es importante subrayar que estas ventajas técnicas adquieren su verdadero valor cuando se integran en un marco interpretativo sólido, articulado desde perspectivas teóricas críticas sobre la violencia de género. La tecnología, en este contexto, no reemplaza la reflexión teórica ni el análisis cualitativo en profundidad; más bien, potencia su alcance al proporcionar bases empíricas más robustas y representativas. El desafío metodológico consiste, precisamente, en articular de manera efectiva las potencialidades técnicas del scraping con los marcos interpretativos que orientan la investigación sobre femicidios.

ENFOQUE HÍBRIDO: INTEGRACIÓN METODOLÓGICA PARA LA INVESTIGACIÓN DE FEMICIDIOS

La propuesta de un enfoque híbrido surge de la necesidad de articular las fortalezas específicas de cada metodología desarrollada respectivamente en las etapas 1 y 2, superando sus limitaciones individuales mediante una integración sistemática que potencie la rigurosidad científica y la profundidad analítica.

Así, el enfoque híbrido que propongo no representa solo una yuxtaposición de técnicas, sino una síntesis metodológica para la triangulación de datos cualitativos y cuantitativos. Esta integración permite aprovechar la capacidad de procesamiento masivo del *web scraping* mientras conserva la profundidad interpretativa y la sensibilidad contextual de las lecturas cualitativas tradicionales.

En una primera etapa, el enfoque híbrido se inicia con la implementación del *web scraping* como herramienta de recolección masiva de datos. Durante esta fase, configuro los *scrapers* para extraer información de medios digitales, estableciendo criterios amplios de búsqueda que capturen el mayor volumen posible de noticias potencialmente relevantes. Los algoritmos de filtrado inicial incluyen no solo términos explícitos como “femicidio” o “violencia de género”, sino también expresiones eufemísticas y descriptivas frecuentemente utilizadas en el tratamiento mediático de estos casos, tal como lo desarrollé en las etapas 1 y 2.

Una segunda fase involucra la revisión manual sistemática de los casos identificados automáticamente. Durante este proceso, como aclaré en la etapa 2 –cuarto procedimiento–, la información

es estructurada en una base de datos que permite su organización y análisis.

Finalmente, realizo una validación cruzada; es decir, para verificar la solidez de la información recopilada, contrasto los datos provenientes de distintas fuentes sobre un mismo caso. Para ello aplico criterios etnográficos para evaluar la pertinencia de cada registro, considerando no solo la precisión de la clasificación automática sino también elementos contextuales nativos que, en muchos casos, es posible determinar en el marco de la interpretación socioantropológica. Entre estos elementos se encuentran el tratamiento mediático del caso, la revictimización, la construcción e interpretación narrativa de la violencia, las representaciones generizadas sobre víctimas, victimarios y sus familiares, y los marcos interpretativos atravesados –o no– por la perspectiva de género por parte de los medios de comunicación, entre otros. Sin este componente analítico, el desarrollo posterior probablemente se reduciría a un procesamiento sofisticado de información superficial.

Lo ideal sería incluir mecanismos de actualización continua que combinen la automatización con la supervisión etnográfica. Aunque aún no lo he establecido, en un futuro proyecto la programación de los *scrapers* para ejecutarse periódicamente, generando alertas sobre nuevos casos.

La fase automatizada de recolección, realizada mediante *web scraping*, constituye el primer paso del enfoque híbrido. Permite reunir grandes volúmenes de información periodística en poco tiempo, agilizando ampliamente la pesquisa. Sin embargo, los resultados del *web scraping* del citado diario *El Tribuno de Jujuy* evidencian los límites de esta etapa.

Por ejemplo, la nota “Fue condenado a perpetua por el crimen de su expareja” (*El Tribuno de Jujuy*, 25 de mayo de 2024) describe el asesinato de Jorgelina Carmen Cruz a manos de su expareja, quien violó una restricción judicial y la atacó a puñaladas. Aunque el cuerpo del texto indica que el homicidio fue calificado “por mediar violencia de género”, el titular evita el término *femicidio*, sustituyéndolo por “crimen”, una forma que diluye el carácter patriarcal del hecho.

Algo similar ocurre en la nota “Hombre asesinó a puñaladas a su expareja y fue detenido” (*El Tribuno de Jujuy*, 13 de octubre

de 2024), donde se informa que un hombre mató a su expareja en la vía pública. La nota se concentra en el procedimiento policial y el operativo de detención, sin contextualizar el vínculo como parte de un patrón de violencia de género. El relato periodístico reduce, en su mayoría, el hecho a un suceso policial indistinguible.

Por su parte, la nota “Prisión perpetua para un hombre que mató a su pareja” (*El Tribuno de Jujuy*, 25 de octubre de 2024) contiene en su desarrollo información fundamental –el reconocimiento judicial del agravante por violencia de género–, pero el titular vuelve a omitir el término *femicidio*.

Como se observa, la automatización agiliza la búsqueda y provee un mapa general, pero es la etnografía la que permite leer los pliegues, las omisiones y los sesgos simbólicos de la violencia femicida narrada. Esto, incluso considerando que *El Tribuno de Jujuy* es uno de los medios locales que ha ido elaborando, a lo largo del tiempo, una narrativa con mayor perspectiva de género en sus titulares.

Por lo tanto, la supervisión etnográfica constituye la instancia que corrige y completa el proceso automatizado. Implica la curaduría de cada nota recuperada, la identificación de prejuicios, eufemismos y omisiones en los textos y, finalmente, la decisión interpretativa sobre el femicidio. En esta etapa, las acciones incluyen revisar cada artículo, analizar su estructura discursiva, el vocabulario y las expresiones empleadas. Así, la supervisión etnográfica transforma un conjunto de datos en una trama de sentidos, haciendo visible lo que el algoritmo no alcanza a leer más allá de los *tokens*.

Esta práctica no es solo correctiva sino profundamente reflexiva. Intenta retomar lo que Rosana Guber explica sobre la destreza antropológica en lo intersticial: esa capacidad de observar lo no documentado, lo implícito y lo no formalizado, de descubrir los intersticios, desfasajes y contradicciones internas (Guber 2005). En este sentido, la etnografía opera sobre los modos en que los medios reproducen el mismo fenómeno patriarcal que sostiene la violencia femicida. El análisis de esos intersticios permite reconocer qué narrativas prevalecen, cuáles se ocultan y cómo los algoritmos contribuyen, indefectiblemente, a esa invisibilización.

En consecuencia, la supervisión etnográfica no puede automatizarse. Su valor reside en la interpretación situada, en la sensibilidad

–que se aprende en el propio campo de estudio– para identificar el sesgo o la omisión y en la capacidad de contextualizar el dato dentro de una trama social más amplia. Es precisamente en esa observancia profunda y concienzuda de los datos donde el enfoque híbrido encuentra su potencia crítica: en articular la eficiencia técnica del *web scraping* con la mirada etnográfica capaz de leer entre líneas la violencia femicida.

CONCLUSIONES

La implementación de un enfoque híbrido para la recolección y análisis de datos sobre femicidios en la provincia de Jujuy representa una innovación metodológica significativa en el campo de los estudios de género y violencia. El recorrido metodológico presentado en este artículo, desde la recolección manual hacia la automatización digital mediante técnicas de *web scraping*, documenta no solo una evolución técnica, sino también una respuesta a los problemas estructurales en la documentación de la violencia femicida en Argentina.

La fragmentación y disparidad de los registros existentes, tanto los gubernamentales como los elaborados por organizaciones de la sociedad civil, constituyen un obstáculo fundamental para la comprensión integral del fenómeno: difieren en definiciones, en el alcance de las categorías y en los criterios de inclusión de los casos, lo que genera resultados fragmentados y difícilmente comparables entre sí. Estas inconsistencias, lejos de ser meramente técnicas, reflejan las tensiones conceptuales y políticas subyacentes en la definición, identificación y contabilización de los femicidios. En este marco, la metodología manual implementada inicialmente, aunque exhaustiva, develó las limitaciones inherentes a los procesos manuales en términos de tiempo, recursos y capacidad de procesamiento.

La prensa, por su parte, tampoco constituye una fuente neutra o completa: en muchos casos omite información, reproduce estereotipos de género o recurre a eufemismos como “crimen pasional” en lugar de *femicidio*. Sin embargo, la propuesta metodológica aquí presentada constituye una mejora frente al *statu quo* porque no se limita a una única fuente ni a una única lógica de registro, sino que articula el potencial del *scraping* automatizado con la supervisión etnográfica crítica.

En primer lugar, el uso de un *scraper* permite ampliar la cobertura y reducir el sesgo de accesibilidad; es decir, se pueden procesar de manera sistemática grandes volúmenes de artículos en plazos breves, superando la lentitud y el subregistro de la búsqueda manual. En segundo lugar, la supervisión etnográfica habilita una lectura situada que detecta los modos en que los medios nombran u omiten la violencia de género, lo cual corrige el posible sesgo de la extracción automatizada. Finalmente, la hibridación de ambas estrategias constituye una mejora frente a los registros existentes, porque permitiría crear series más completas, replicables y críticas, al tiempo que hace visible cómo se producen las narrativas sobre los femicidios en su propio contexto. En suma, la mejora no radica en suponer que la prensa sea más consistente que otras fuentes, sino en combinar eficiencia técnica con interpretación crítica, reduciendo los vacíos y las inconsistencias que caracterizan, en su mayoría, al actual panorama de registros.

La transición hacia el *web scraping* mediante Python emerge como una respuesta a estas limitaciones, ofreciendo ventajas significativas en términos de eficiencia, alcance y sistematización. Sin embargo, resulta fundamental comprender que esta evolución metodológica no implica un abandono de la perspectiva cualitativa, sino una complementación que potencie ambos enfoques. La automatización en la recopilación de datos libera recursos para profundizar en el análisis interpretativo, permitiendo una aproximación más compleja al fenómeno estudiado.

Este enfoque híbrido se inscribe en las discusiones contemporáneas sobre métodos digitales en la investigación social. La integración de herramientas computacionales con perspectivas críticas sobre la violencia de género permite superar tanto las limitaciones técnicas de los métodos manuales como los sesgos potenciales de un enfoque exclusivamente cuantitativo o automatizado.

La experiencia desarrollada en la provincia de Jujuy, caracterizada por alarmantes índices de femicidio en el contexto nacional, demuestra el potencial de esta metodología para la construcción de registros más completos y representativos. Los resultados preliminares, que identifican un número significativamente mayor de víctimas que los registros oficiales, evidencian el subregistro

existente y refuerzan la necesidad de implementar aproximaciones metodológicas innovadoras y rigurosas.

El camino recorrido hasta aquí abre nuevas perspectivas para la investigación sobre violencia de género, proyectando la posibilidad de extender este enfoque a otros contextos regionales y temporales. Asimismo, establece las bases para el desarrollo de un protocolo metodológico integral que articule efectivamente las potencialidades técnicas del *scraping* con marcos interpretativos críticos sobre los femicidios.

En síntesis, la transformación metodológica expuesta en este trabajo trasciende su dimensión técnica para configurarse como una contribución significativa a la visibilización y comprensión de la violencia femicida. La articulación entre perspectivas tradicionales y herramientas digitales emerge como una estrategia promisorio no solo para la construcción de conocimiento académico, sino también para la fundamentación de políticas públicas orientadas a la prevención y erradicación de esta forma extrema de violencia contra las mujeres.

REFERENCIAS BIBLIOGRÁFICAS

- Almérás, Diana, y Coral Calderón, coords. 2012. *Si no se cuenta, no cuenta: información sobre la violencia contra las mujeres*. Santiago de Chile: CEPAL.
- Amnistía Internacional. 2008. *Muy tarde, muy poco. Mujeres desprotegidas ante la violencia argentina*. Buenos Aires: Amnistía Internacional Argentina.
- Asociación Civil La Casa del Encuentro. 2020. *Por Ellas: 10 años de informes de femicidios en Argentina*. Ciudad Autónoma de Buenos Aires.
- Barcaglione, Gabriela. 2005. “Femicidios: cómo los medios construyen noticias”. En *Femicidios e impunidad*, editado por Silvia Chejter. Argentina: CECYM. https://americalatinagenera.org/wp-content/uploads/2014/06/doc_367_FemicidioseImpunidad.pdf.
- Benítez, Lucía. 2019. “Ética y transparencia para la detección de sesgos algorítmicos de género”. *Estudios sobre el Mensaje Periodístico* 25, 3: 1307–1320.
- Caputi, Jane, y Diana Russell. 1990. “Femicide: Speaking the Unspeakable”. *The World of Women* 1: 34–37.
- Corte Suprema de Justicia de la Nación. 2016. *Registro Nacional de Femicidios en la Justicia Argentina*. Argentina: Oficina de la Mujer, Corte Suprema

- de Justicia de la Nación. <https://om.csjn.gov.ar/consultaTalleresWeb/public/documentoConsulta/verDocumentoById?idDocumento=166>.
- Corte Suprema de Justicia de la Nación. s.f. *Metodología: Registro Nacional de Femicidios en la Justicia Argentina*. Argentina: Oficina de la Mujer, Corte Suprema de Justicia de la Nación. Consultado el 26 de septiembre de 2025. <https://www.csjn.gov.ar/omrecopilacion/omfemicidio/metodologia.html>.
- Defensoría General de la Nación. 2015. *Femicidio y debida diligencia: estándares internacionales y prácticas locales*. Ciudad Autónoma de Buenos Aires: Amnistía Internacional.
- Defensoría General de la Nación. 2018. *Informe Final. 1º de enero al 31 de diciembre de 2017. Observatorio de Femicidios*. https://www.dpn.gob.ar/documentos/Observatorio_Femicidios_-_Informe_Final_2017.pdf.
- D'Ignazio, Catherine, Helena Suárez, Silvana Fumega, Harini Suresh, Isadora Cruxên, Winnie So, Andrea Martinez, y Mariel García-Montes. 2020. *Feminicide and Machine Learning: Detecting Gender-based Violence to Strengthen Civil Sector Activism*. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/60535>.
- El Tribuno de Jujuy. 2024. "Fue condenado a perpetua por el crimen de su expareja". 25 de mayo. <https://eltribunodejujuy.com/policiales/2024-5-25-1-0-0-fue-condenado-a-perpetua-por-el-crimen-de-su-expareja>.
- El Tribuno de Jujuy. 2024. "Hombre asesinó a puñaladas a su expareja y fue detenido". 13 de octubre. <https://eltribunodejujuy.com/policiales/2024-10-13-0-0-0-hombre-asesino-a-punaladas-a-su-expareja-y-fue-detenido>.
- El Tribuno de Jujuy. 2024. "Prisión perpetua para un hombre que mató a su pareja". 25 de octubre. <https://eltribunodejujuy.com/policiales/2024-10-25-0-25-0-prision-perpetua-para-un-hombre-que-mato-a-su-pareja>.
- Gambetta, Victoria. 2022. "Dificultades y desafíos para investigar el femicidio en Latinoamérica". *Revista Latinoamericana de Metodología de las Ciencias Sociales* 12 (2).
- Gómez, Carina Elizabeth. 2020. "Retratos de Memorias: análisis de fotografías documentales de familiares de víctimas de femicidio (Jujuy-Argentina)". *Iberoamérica Social. Revista-Red de Estudios Sociales* XIV: 116–139. <https://ri.conicet.gov.ar/handle/11336/144593>.

- Gómez, Carina Elizabeth. 2022. “Es mentira que el tiempo pasa. El tiempo se atora’. Reflexiones acerca de una investigación antropológica en torno a la violencia femicida”. *Narrativas Antropológicas*, ENAH. <https://www.revistas.inah.gob.mx/index.php/narrativasantropologicas/article/view/18742/20086>.
- Jayachandran, Jayasankar, y Vijay Arni. 2023. *Traversing the Ethical Landscape of Data Scraping for ai*. SSRN. <https://ssrn.com/abstract=4666354>.
- La Casa del Encuentro. s.f. *Informe de investigación de femicidios en Argentina*. Consultado el 26 de septiembre de 2025. <https://www.lacasadelencuentro.org/femicidios03.html>.
- Luscombe, Alex, Kevin Dick, y Kevin Walby. 2022. “Algorithmic Thinking in the Public Interest: Navigating Technical, Legal, and Ethical Hurdles to Web Scraping in the Social Sciences”. *Qualitative and Quantitative Methods* 56: 1023–1044.
- Ministerio de las Mujeres, Género y Diversidad. 2023. *Encuesta de prevalencia de violencia contra las mujeres*. Argentina: Ministerio de las Mujeres, Género y Diversidad.
- Mujeres de la Matria Latinoamericana (MuMaLa). 2020. *Registro nacional de Femicidios, Femicidios Vinculados, Trans/Travesticidios y Lesbicidios de MuMaLa. Año 2020. Parcial y resumen en contexto pandemia covid-19*. <https://libresdelsur.org.ar/wp-content/uploads/Registro-nacional-femicidios-de-MuMaLa-2020.pdf>.
- Mujeres de la Matria Latinoamericana (MuMaLa). 2022. *7 años del #NiUnaMenos. Registro nacional de Femicidios, Femicidios Vinculados, Trans/Travesticidios de MuMaLa. Año 2015/mayo 2022*. Argentina: Observatorio Nacional MuMaLa.
- Ni Una Menos. 2015. “Manifiesto #1”. 3 de junio. <http://niunamenos.org.ar/manifiestos/3-de-junio-2015/>.
- Observatorio de Femicidios de la Defensoría del Pueblo de la Nación. 2020. *Femicidios. Informe Anual 2020*. Buenos Aires: Defensoría del Pueblo de la Nación Argentina.
- Página 12. 2012. “El femicidio fue mayoría”. 2 de diciembre. <http://www.pagina12.com.ar/diario/sociedad/3-209040-2012-12-02.html>.
- Redacción Clarín. 2008. “La violencia de género provocó 240 muertes en lo que va del año”. *Clarín*, 25 de noviembre. https://www.clarin.com/ultimo-momento/violencia-genero-provoco-240-muertes-va-ano_o_BipxJLoRTtx.html.

- Rico, Ada Beatriz, y Francisco Tuñez. 2013. *Por Ellas... 5 años de Informes de Femicidios. Observatorio de Femicidios en Argentina 'Adriana Marisel Zambrano' de la Asociación Civil La Casa del Encuentro*. Buenos Aires: La Casa del Encuentro.
- Suárez, Helena. 2020. "Datos Discordantes: Información Pública Sobre Femicidio en Uruguay". *Mundos Plurales. Revista Latinoamericana de Políticas y Acción Pública* 7, 1: 53–78.
- Vilkova, Olga. 2020. "Web Scraping as a Method of Data Extraction in Sociological Studies: On Scientific Applicability". *Vestnik Tomskogo Gosudarstvennogo Universiteta. Filosofiya, Sotsiologiya, Politologiya* 54: 163–175.