

Un abordaje computacional a las protestas de los trabajadores de la salud pública de Chaco entre 2017 y 2023

Julio Andrés Sujatt (UNNE – IIGHI // UNLP)

Este artículo tiene dos objetivos. Por un lado busco presentar una mirada de los conflictos que hayan tenido a trabajadores de la salud como protagonistas durante los últimos cinco años en el sistema de salud pública de la Provincia de Chaco. A su vez propongo comentar los procesos metodológicos de recolección, procesamiento, presentación y análisis de datos trabajados con herramientas computacionales de programación. Esta modalidad de trabajo en las ciencias sociales aún sigue siendo una novedad y lo que se presenta en el artículo es, de hecho, el resultado de una primera experiencia con la programación. Es por esto que también presento algunas recomendaciones, reflexiones y conclusiones pensando en contribuir a la creciente adopción de estas herramientas.

Investigar usando lenguaje de programación

Durante el período abordado los trabajadores de la salud pública de Chaco – profesionales y no profesionales, de planta permanente o contratados de manera precaria, sindicalizados y no sindicalizados – desplegaron de manera recurrente acciones de protesta para sostener sus demandas. El contexto se ha mostrado cambiante en base a dos eventos relevantes para las dinámicas de protesta del sector. En 2019 hubo elecciones que culminaron con el recambio de las máximas autoridades de los gobiernos nacional y provincial. El otro evento trascendente ha sido la pandemia de covid-19 iniciada en 2020 con un fuerte impacto en las condiciones laborales del sector salud que ya se encontraban deterioradas.

Para la investigación del fenómeno de las acciones de protesta en la salud pública de Chaco se utilizaron como fuentes de información las notas periodísticas del portal web de noticias *Diario Norte*, de Resistencia, en su sección Locales. Desde el punto de vista computacional estas notas en la web componen un conjunto de datos semi-estructurados que deben ser extraídos y convertidos a un formato de datos estructurados -organizados de manera unificada, homogénea y tabular- para poder trabajar, por lo que estos procedimientos son los primeros pasos en el trabajo computacional.

Los procedimientos metodológicos computacionales se organizan y ejecutan a partir de una secuencia lógica de funciones codificadas, a las que se llama *scrip*, y escritas en líneas de códigos con un lenguaje de programación. Los códigos de estas funciones se agrupan en librerías o paquetes que componen el acervo de herramientas del lenguaje de programación. En este sentido cada código establece una tarea o función en el marco de un encadenamiento ordenado de tareas orientadas a cumplir un objetivo y se ejecutan en un entorno computacional adecuado. Aquí se trabajó con el lenguaje de programación R con el *software* de entorno de desarrollo integrado para escritorio RStudio ¹. En términos de hardware se utilizó una PC de escritorio hogareña con conexión a internet. De todas maneras Nieto asegura que puede optarse por opciones más baratas y reducidas (2021: 127).

A modo de ejemplo de *scrip* en la siguiente imagen pueden verse algunas de las primeras líneas de código en una parte del entorno RStudio donde se programan el acceso a la página web (línea 14), la creación de un vector con la secuencia de las páginas que contienen las notas de la sección Locales (línea 16), el ingreso a cada una de esas páginas para capturar el link de cada nota y agruparlos en un nuevo vector (líneas 18 a 24) y finalmente depurar esa lista de campos vacíos y guardarla (líneas 26-27).

Imagen I: líneas de código del scrip desarrollado.

```
14 url <- 'https://www.diarionorte.com/notas/'
15
16 urls <- c(paste0(url,'locales/?page=',2:1725))
17
18 links <- c()
19 for(url in urls[1:1720]) {
20   html <- read_html(url)
21   links <- c(links, html %>%
22     html_nodes(xpath='//*[contains(@class, "grid")]/article/a') %>%
23     html_attr("href") %>%
24     url_absolute(url)) }
25
26 links <- na.omit(links)
27 saveRDS(links, "links.rds")
28
29
```

Elaboración propia.

Para la elaboración del *scrip* se buscó seguir el esquema desarrollado por Laitano y Nieto (2022) y con base en las líneas de códigos presentadas en los *scripts* del seminario “Herramientas computacionales para el análisis de la conflictividad social” (Agustín Nieto – UNLP – 2022). Como herramienta de consulta externa frente a complicaciones en la adaptación de las líneas de código se recurrió a ChatGPT de OpenAI.

1 R es un lenguaje de programación de código libre: <https://www.r-project.org/>
RStudio está disponible en una versión de código libre para su descarga gratuita: <https://posit.co/>

Laitano y Nieto (2022: 156 - 157) señalan cuatro pasos para un trabajo más completo pero similar al aquí expuesto. El primer paso, ya señalado en parte, es llamado *web scraping* y consiste en la recolección en masa y automática de información semi-estructurada en internet y su pasaje a un formato de tabla. “Cada fila de la tabla refiere a una única nota publicada (...), mientras que cada columna refiere a una parte de la nota como puede ser el título, o el subtítulo o entradilla, el cuerpo de la nota, la fecha, el link, la foto, etc.” (Nieto, 2021: 126). Para esto se utilizó la librería *rvest*. Las segunda y tercera etapa refieren a herramientas de *text mining* que implican explorar, analizar y modelar grandes cantidades de datos con el objetivo de extraer información relevante. En este trabajo se usaron técnicas con diccionarios para la clasificación y cuantificación de notas relevantes. El análisis de los datos requiere de paquetes agrupados en *tidyverse* junto con *ACEP*. Para la cuarta etapa, de visualización, se utilizaron funciones base de R como *barplot*. La importancia de poder visualizar datos es que “continúa el análisis exploratorio y avanza en el testeo de hipótesis” (Nieto, 2021: 126).

Para el estudio de fenómenos como el de la recurrencia de acciones de protesta o de conflictos el salto de productividad que ofrece esta herramienta es evidente frente a la inspección y registro ‘artesanal’ en fuentes periodísticas. Sin embargo hay restricciones. La primera y más evidente es que el *scraping* no puede suplantar al método artesanal si la fuente no está enteramente disponible en la web². Por caso el objetivo original para este trabajo era *scrapear* desde el año 2015 los conflictos del sector salud chaqueño. Sin embargo ninguno de los portales de prensa considerados para esta exploración daba acceso a notas de los años 2015 y 2016 en las secciones requeridas, siendo el más completo Diario Norte, con noticias desde enero de 2017.

Una vez definida la fuente mediante *scraping* se extrajeron titulares, fechas, volantas, bajadas y cuerpos de las notas. La inspección de la página tuvo cierta dificultad sobre todo para extraer las fechas correspondientes, lo que implica identificar la sección específica en el código de diseño de la página teniendo en cuenta que todas las notas cuentan con una estructura de diseño similar. La solución se construyó, en este y en los siguientes obstáculos, con prueba y error y usando como herramienta externa ChatGPT. Fue interesante usar esa herramienta novedosa ya que, según se le podía, ofrecía numerosas opciones para armar códigos. La mayoría de las veces dichos códigos contenían errores imprevistos o tenían una estructura muy diferente a las que se venía

2 Existen herramientas más complejas que permiten convertir imágenes a texto (como por ejemplo notas de diarios escaneadas a texto estructurado).

usando – ante lo que se buscó mantener el *scrip* dentro de las funciones ya conocidas –, sin embargo ChatGPT sirvió efectivamente para superar varias dificultades de manera ‘dialogada’.

El *scraping* también tuvo dificultades ‘técnicas’. Debido a una mala conexión a internet fueron varios los intentos frustrados y los esfuerzos por continuar la extracción de datos desde el punto interrumpido, sin perder el trabajo previo. De acuerdo al trabajo requerido y al protocolo utilizado el *scrapeo* puede durar varias horas³.

Una vez obtenido y estructurado el corpus de notas en una tabla se procedió al uso de ‘diccionarios’. Esta técnica consiste en discriminar notas de acuerdo a la identificación de palabras clave. El corpus original se componía de 19.647 notas completas. Con la aplicación del diccionario sobre todos los segmentos de las notas se seleccionaron 10.970 que contenían al menos una de las palabras del diccionario SISMOS del paquete *ACEP*. Sin embargo la utilización de este diccionario pre-configurado también tuvo su complicación. El diccionario contenía al momento de consulta 219 palabras destinadas a detectar eventos de protesta en cuerpos de texto incluyendo la entrada ‘resisten’. Orientada claramente a capturar notas sobre eventos o discursos de carácter contestatarios y/o defensivos - resistencia-, en el marco chaqueño esta palabra generalmente refiere a la ciudad capital y su área metropolitana. Es por esto que ‘resisten’ tuvo que ser suprimida para evitar falsos positivos. Para dar una idea de la magnitud de los falsos positivos que imputaba esta palabra, haciendo correr el diccionario SISMOS solo entre los títulos de las notas se capturaron 3386 positivos que luego de suprimir la entrada ‘resisten’ del diccionario se redujeron a 2244. Casi un 34% menos de títulos que tenían a ‘resistencia’ como única coincidencia con el diccionario. Conociendo el uso habitual de la palabra en la prensa escrita local y a falta de otra coincidencia relevante para eventos de protesta podemos suponer que eran falsos positivos que hacían referencia a la ciudad y no a acciones de protesta. En definitiva manteniendo la entrada ‘resisten’ el *corpus* total de notas solo se hubiera reducido a 14.887 observaciones.

El *corpus* de 10.970 notas, no obstante, resultaba excesivo por ser inverosímil que casi el 56% del total de las notas de la sección Locales refirieran a un conflicto relevante. Por ello ese *corpus* fue nuevamente procesado con un diccionario *ad-hoc* compuesto

3 Debe tenerse en cuenta que un uso inadecuado de esta técnica puede provocar el colapso en algunas páginas web no preparadas para soportar gran cantidad de visitas. Existen paquetes para administrar esto.

solo por una docena de palabras que hicieran referencias a conflictos en el sector de salud pública chaqueño. Con una primera versión de este breve diccionario la base de datos se redujo a 3.987 notas. Si bien la reducción fue notoria se volvió a ajustar el diccionario a solo seis palabras, eliminando entradas genéricas como ‘salud pública’ y ‘hospital’ para solo dejar términos específicos que identificaran a los actores que interesan: ‘residente’, ‘aptasch’, ‘ate’, ‘gremial medica’, ‘upcp’, ‘programa experto’.⁴ Resultaron 548 observaciones.

En el diseño del diccionario el contenido que va entre comillas debe adaptarse lo mejor posible a lo que se desea encontrar haciendo uso, por caso, de los espacios para abrir y cerrar explícitamente los términos o escribiendo solo la raíz de las palabras para capturar sus variantes. Por ejemplo con ‘residente’ pretendemos capturar las notas que hagan referencia a los médicos que se desempeñan en los hospitales bajo el particular régimen académico/laboral de residencia. El espacio inicial dejado tras la primera comilla busca que solo se capturen palabras que inician exactamente como la marcada. De otra manera el buscador captura el componente ‘residente’ indistintamente, ya sea como palabra única o como parte de una palabra mayor, como por ejemplo ‘presidente’. Sin embargo no se puede evitar que el código capture referencias a notas que nombran como ‘residentes’ a vecinos de barrios. Algo similar sucede con ‘ate’. Para que el código solo capture referencias a la organización sindical (Asociación de Trabajadores del Estado) dejamos espacios luego de la comilla de apertura y antes de la de cierre para que no se encuentren coincidencias con cualquier otra palabra que contengan esa combinación de letras, como ‘debate’ ‘patente’.

La conflictividad en el sector de la salud pública de Chaco

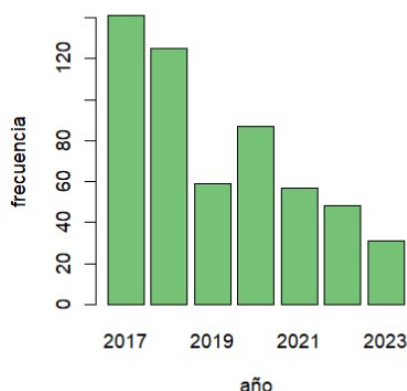
Con estos procedimientos se logró estructurar una base de datos que organice el *corpus* de fuentes periodísticas relevantes sobre los eventos de protesta en el sector sanitario de

4 ‘residente’: refiere a graduados universitarios de las distintas disciplinas de la salud que cumplen residencia en hospitales como parte de un programa de formación académica y profesional de posgrado.
‘aptasch’: Asociación de profesionales, técnicos y auxiliares de la salud de la provincia de Chaco, organización sindical con importante representación específica en el sector.
‘ate’: Asociación de trabajadores del Estado, organización sindical con importante representación en casi todos los organismos del Estado, incluyendo los de salud.
‘gremial medica’: Gremial Médica Hospitalaria de Chaco, organización sindical que nuclea a profesionales de la medicina
‘upcp’: Unión de personal civil de la Provincia del Chaco, organización sindical más importante de la provincia, con importante representación en el sector.
‘programa experto’: El Programa Experto abarca en la provincia de Chaco a trabajadores de la salud pública con modalidades precarias de contratación y de salario.

la región en los últimos años. A su vez la frecuencia de notas (Nieto, 2021: 160) con menciones a conflictos ofrece una referencia indirecta a la distribución temporal de hechos de este tipo en la región. A este nivel de análisis es apresurado suponer que a cada nota sobre eventos de protesta le corresponde efectivamente un conflicto, sin embargo esta referencia ofrece una aproximación inicial válida sobre la magnitud de la conflictividad.

El gráfico presentado a continuación permite realizar algunas observaciones.

Gráfico I: notas con referencias a conflictos en el sector salud por año. Chaco. (Nota: El año 2023 contempla hasta el mes de abril).



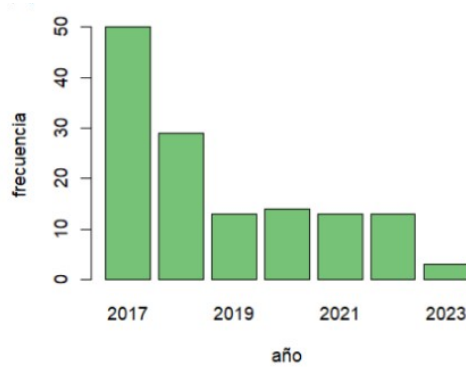
Fuente: Diario Norte. Elaboración propia.

Se muestra la distribución de notas que expresan eventos de protesta en el sector de salud y se observa que los años 2017 y 2018 tuvieron una base alta de conflictividad que en los años siguientes no volvió a aproximarse ni siquiera en el ‘repunte’ 2020. Se puede hipotetizar que la nueva ‘normalidad’ pos 2019 refiere a un nuevo escenario político en la provincia donde los conflictos son contenidos o se canalizan de manera más institucional o política. Que la tendencia haya iniciado en 2019 (último año del gobierno de Domingo Peppo) puede explicarse por las expectativas electorales de ese año en los que las declaraciones y movilizaciones políticas pudieron haber ocupado el lugar de canalizar los malestares y las demandas que habitualmente se expresan sindicalmente. Teniendo en cuenta que abordamos un fenómeno social con la mediación de fuentes periodísticas de una empresa de la región también debe considerarse la posibilidad de la aparición de un nuevo sesgo, ya sea en el uso de palabras o en el tratamiento periodístico de los eventos, pero esto requiere un análisis diferente que excede a este trabajo.

Debe mencionarse que a pesar de haber pasado varios filtros, en un examen manual del corpus ⁵ se detectaron varios ‘falsos positivos’. En este punto se tomó la decisión de realizar un recorte más enfocado para trabajar con notas que se refieran al conflicto en el sector salud como tema principal y el *corpus* fue nuevamente tamizado con el diccionario *ad hoc* para seleccionar solo las que contuvieran las palabras clave en sus títulos o bajadas. Surgió un cuerpo reducido de 135 notas. Esta decisión implica el riesgo de dejar de lado información relevante contenida en el resto de las notas y a su vez mantiene el inconveniente de que palabras como ‘ate’ y ‘upcp’ continúen capturando notas que se refieren a conflictos que no son específicos del sector de salud debido a que ambas son organizaciones que operan en el resto de las dependencias del Estado provincial. Debido a estas salvedades este ejercicio es ilustrativo y sus conclusiones no son definitivas.

El siguiente gráfico muestra la nueva forma de distribución de las notas con referencias a conflictos en el sector salud en base a sus títulos y bajadas. En términos generales se mantienen las tendencias que habíamos observado en base al *corpus* más amplio de notas, pero aquí sobresale aún más el pico de 2017 y se atenúa el repunte de 2020 para ser casi imperceptible.

Gráfico II: notas con referencias a conflictos en el sector salud en titular o bajada por año. Chaco. (Nota: El año 2023 contempla hasta el mes de abril).



Fuente: Diario Norte. Elaboración propia.

En los siguientes cuadros se avanza en el contenido de las notas en lo referente a la modalidad de protestas, las demandas y los actores que las protagonizan.

Tabla I. Notas con menciones a acciones de protesta en título o bajada, por modalidad y año en Chaco.

Año	Paros/ huelgas	Movilizaciones
-----	----------------	----------------

5 Con la función `datatable` del paquete `DT` se pueden generar tablas dinámicas para el uso grandes bases de datos. Pueden exportarse, por ejemplo, al formato `html` para ser usadas cómodamente con cualquier navegador web.

2017	35	22
2018	13	16
2019	7	8
2020	5	5
2021	3	5
2022	6	5
2023*	3	0

Fuente: Diario Norte. Elaboración propia. * Enero – abril de 2023

Tabla II. Notas con menciones a demandas en título o bajada, por tipo y año en Chaco.

Año	Demanda salarial	Demanda por precarización
2017	27	16
2018	19	5
2019	4	6
2020	8	5
2021	6	5
2022	7	4
2023*	0	1

Fuente: Diario Norte. Elaboración propia. * Enero – abril de 2023

Tabla III. Notas con menciones a sujeto protagonista en título o bajada, por año en Chaco.

Año	Sindicalismo	Trabajadores Prog. Expertos	Residentes de salud
2017	45	1	3
2018	28	1	0
2019	10	0	1
2020	10	2	4
2021	5	4	2
2022	7	0	5
2023*	2	0	0

Fuente: Diario Norte. Elaboración propia. * Enero – abril de 2023

Con el objetivo de describir la modalidad prevaleciente de herramientas de protesta de cada año en el sector salud, para la elaboración del cuadro I se contabilizaron por año la cantidad de notas que hicieran mención a acciones de paro o de movilización. Se eligieron este tipo de acciones por considerarlas, *a priori*, las más representativas del repertorio de lucha del movimiento de trabajadores. Se aplicó un diccionario *ad hoc* con

las palabras ' paro', ' paran', ' huelga' y, a continuación, se aplicó otro con las palabras ' moviliz' y ' manifest'. Posteriormente se contabilizaron las notas con referencias positivas y se las organizó en formato tabla. Los mismos procedimientos se realizaron para los cuadros II y III para identificar tipos de demandas significativas y actores específicos.

Para el cuadro II se usaron las palabras ' aumento', ' salari', ' suma fija' y ' precariz' y ' pase a planta'. El objetivo en este caso fue evaluar el peso de cada agenda de demandas: la salarial, común a todos los trabajadores, o la de precarización acentuada en el sector sometido a condiciones desfavorables de contratación. En el cuadro III se buscó capturar notas con referencias a los actores sindicales con actuación en el área de salud mediante las palabras ' aptasch', ' ate ', ' gremial', ' upcp' y ' sindica'. El universo se completó con búsquedas para 'programa experto' y ' residente'. De esta manera se buscó sopesar la presencia de estos actores por separado.

Se evidencia una notable reducción en las magnitudes a partir del año 2019 tanto para notas referentes a acciones de protesta como para las que se refieren a demandas. Sin embargo hay un interesante fenómeno referente a colectivos de protesta no sindicalizados: los precarizados del Programa Expertos y los médicos residentes. Estos tomaron presencia en los años 2020 y 2021 y los residentes, además, continuaron con presencia fuerte en 2022. Para corroborar esta tendencia en el marco de la fuente utilizada, en la siguiente tabla se observan los mismos fenómenos utilizando las mismas herramientas de análisis sobre el corpus de 548 observaciones con menciones a conflictos vinculados al sector de salud pública en todo el cuerpo del texto. Los protagonistas precarizados del sistema (residentes y 'expertos') mostraron un crecimiento exponencial de su presencia en las notas con referencias a protestas. Mientras en 2017 y 2018 representaban un 9,6% y un 8,4% de total respectivamente, en 2019 representaron un 25%, y llegaron a un pico del 67,5% en 2021 (con 53,3% en 2020 y 47% en 2022). Claramente se confirma una creciente presencia de estos actores a la par que descende la magnitud total de notas con referencias a actores de protesta por año: el año 2020 tiene menos de un 50% de menciones que 2017. Los años 2019 y 2021 están por debajo del tercio.

Esto permite elaborar algunas hipótesis explicativas para continuar investigando. ⁶

6 Una contextualización y un análisis más detallado del fenómeno con base en un enfoque cualitativo se encontrará en Sujatt, Julio Andrés, "Las condiciones de trabajo y la organización de trabajadores precarizados de la salud en hospitales de Resistencia – Chaco". *Revista Conflicto Social*. En prensa.

Tabla I. Notas con menciones a actores protagonistas de protestas por año en todo el cuerpo del texto (enero 2017 – abril 2023).

Año	Sindicales	Trabajadores ‘Programa Expertos’	Residentes de salud	Notas por año
2017	113	3	9	125
2018	98	1	8	107
2019	30	5	5	40
2020	28	5	27	60
2021	13	5	22	40
2022	18	1	15	34
2023	12	0	8	20

Fuente: Diario Norte. Elaboración propia.

Comentarios finales

Las conclusiones, como se mencionó, son provisorias y referidas a este ejercicio de exploración que todavía requiere ajustes. Los datos recolectados permiten suponer que la nueva mediación política que empezó a tomar forma en 2019 tuvo efectos eficientes en reorientar o atemperar las demandas y acciones de protesta de origen sindical en el sector salud. En paralelo la pandemia de covid-19 tuvo efectos agudizantes sobre deficiencias crónicas en las condiciones de trabajo del sector salud. Esta situación parece expresarse en la contraposición entre el retroceso de la presencia del actor sindical como dinamizador de hechos de protesta desde 2019 y la aparición recurrente desde 2020 de actores no institucionalizados en el ámbito del conflicto, los trabajadores precarizados del ‘Programa Expertos’ y los médicos residentes de los hospitales de la provincia, que no obstante no alcanzan a tomar el volumen de los actores sindicalizados. Para hacer una evaluación situada de estos fenómenos descriptos también debe considerarse que en el período abordado la crisis económica argentina se ha agudizado particularmente para los distintos estamentos de la clase trabajadora, con una continuada caída del ingreso y ajustes en las cuentas estatales que impactan en los servicios públicos.

La construcción de una metodología de análisis computacional de eventos de protesta en base a técnicas de *scraping web* y minería de datos permitió, en este caso, mostrar un gran potencial para el procesamiento de grandes cantidades de datos junto con la

flexibilidad para ir ajustando el foco en los aspectos que se consideraron más relevantes a la investigación: actores, modalidades de protesta y demandas.

BIBLIOGRAFÍA

- Laitano, Guillermina y Nieto, Agustín (2022) La conflictividad social en los barrios de Mar del Plata (2016-2020): un acercamiento computacional. En Laitano y Nieto *La conflictividad social en la historia reciente: estudios sociohistóricos entre lo local y lo regional*, Buenos Aires, TeseoPress
- Nieto, Agustín (2021) *Intersecciones entre historia digital e historia social: un ejercicio de lectura distante sobre la conflictividad marítima en la historia argentina reciente*. Drassana, 28, 122–142

FUENTES PERIODÍSTICAS

- Diario Norte. Sección Locales. Enero de 2017 – Abril de 2023. Ver: <https://www.diarionorte.com/notas/locales/>

REPOSITORIO / LIBRERIAS R

- ACEP. Ver: <https://agusnieto77.github.io/ACEP/>
- base
- dplyr
- DT
- lubridate
- rvest
- tidyverse

SOFTWARE

- R / RStudio