



CLIENT TRACKER: GEOLOCALIZACIÓN ABIERTA PARA ANALÍTICA EDITORIAL Y MAPEO DE AUDIENCIAS

Autores: Annier Jesús, Fajardo Quesada¹, René, Herrero Pacheco², Rolando Javier, Álvarez Pérez³, Eduardo Antonio, Hernández González⁴, Jimmy Javier, Calás Torres⁵

¹ Universidad de Ciencias Médicas de Granma. Facultad de Ciencias Médicas de Bayamo. Granma, Cuba. ² Hospital Clínico Quirúrgico Provincial Carlos Manuel de Céspedes. Granma, Cuba. ³ Universidad de Ciencias Médicas de Granma. Facultad de Ciencias Médicas Celia Sánchez Manduley. Granma, Cuba. ⁴ Universidad de Ciencias Médicas de Pinar del Río. Pinar del Río, Cuba. ⁵ Universidad de Ciencias Médicas de la Habana. La Habana, Cuba.

annierfq01@gmail.com

Resumen

Introducción: La medición del alcance territorial de contenidos académicos es clave para la internacionalización. La interrupción de servicios de terceros para geolocalización dejó a muchas revistas científicas cubanas sin una herramienta abierta y reproducible. **Objetivo:** Describir el diseño y desarrollo de una aplicación web de acceso abierto para identificar la localización aproximada de usuarios y visualizarla en mapas interactivos. **Método:** Investigación de desarrollo e innovación tecnológica (abril 2024). La aplicación se diseñó en Ubuntu 24.04 utilizando JavaScript (React) y Python (FastAPI) en Visual Studio Code. Su creación involucró tres etapas: búsqueda de información, selección de herramientas de programación y desarrollo de la aplicación. **Resultados:** El sistema obtiene el país y, cuando es posible, la región/ciudad (con radio de precisión configurable), agrega eventos en tiempo real y visualiza datos mediante mapas de calor y marcadores en una interfaz responsiva. Ofrece integración REST/JS, registro auditado, y opciones de anonimización y retención de datos. **Conclusiones:** ClientTracker proporciona una alternativa ligera y auditable para visualizar el alcance geográfico de contenidos académicos,



cumpliendo con buenas prácticas de privacidad y con rutas de mejora en precisión, rendimiento y gobierno de datos.

INTRODUCCIÓN

La medición del alcance territorial de los contenidos académicos trasciende la mera curiosidad estadística para erigirse como una herramienta estratégica fundamental para las editoriales y portales científicos. Visualizar geográficamente la distribución de lectores y autores proporciona una capa de inteligencia invaluable que informa directamente decisiones críticas de negocio y difusión del conocimiento. Este análisis permite identificar mercados emergentes para la internacionalización, optimizar estrategias de marketing dirigidas, priorizar inversiones en traducción de artículos según la demanda regional y evaluar la idoneidad de índices y bases de datos específicos para la audiencia real de la publicación. Por consiguiente, la integración de mapas de datos en los tableros de mando editoriales se convierte en un componente indispensable para cualquier organización que busque maximizar su impacto global de manera informada y eficiente. (1)

Históricamente, la implementación de esta funcionalidad se ha sustentado en la dependencia de servicios externos. Muchas revistas y portales institucionales incorporaron widgets o plataformas de terceros que, de forma sencilla, generaban mapas de visitas atractivos visualmente. Sin embargo, esta practicidad generó una vulnerabilidad latente. La caída o el cierre repentino de varios de estos servicios, un fenómeno común en el dinámico ecosistema digital, dejó un vacío tanto funcional como metodológico. Este episodio puso de manifiesto una problemática subyacente: si bien el uso de estos mapas estaba ampliamente extendido, existía una escasa estandarización y documentación sobre los procesos de creación de estas herramientas, lo que impedía su reconstrucción o despliegue reproducible de manera autónoma e independiente.

En paralelo, el panorama tecnológico y normativo relacionado con la privacidad y el seguimiento de usuarios ha experimentado una transformación radical y acelerada. La creciente regulación (como el GDPR) y las políticas de los navegadores han llevado a la deprecación progresiva de las cookies de terceros,



elemento central del tracking tradicional. Asimismo, la dirección IP es ahora ampliamente considerada un dato de carácter personal, cuyo tratamiento exige máximas garantías. Este nuevo paradigma exige soluciones técnicas que adopten enfoques server-side, minimicen la recolección de datos y documenten de forma transparente y auditable su tratamiento para garantizar el cumplimiento legal. La antigua dependencia de servicios externos no auditables se vuelve insostenible en este contexto, creando una necesidad apremiante de herramientas alternativas. (2)

La importancia de este desarrollo de un sustituto reside en abordar un problema multifacético: la pérdida de una funcionalidad crítica, la necesidad de soberanía tecnológica y el imperativo de cumplir con los más altos estándares de privacidad. Por lo anterior se desarrolla el presente estudio con el objetivo de describir el diseño y desarrollo de una aplicación web de acceso abierto para identificar la localización aproximada de usuarios y visualizarla en mapas interactivos.

MÉTODO

Clasificación del estudio. Innovación tecnológica (creación de software) con evaluación funcional interna.

El equipo se integró por de 1 desarrollador de software, que laboro desde noviembre hasta diciembre del 2024. Por la simplicidad de la aplicación no se hizo control de versiones en repositorios de git.

Arquitectura y flujo

1. Captura: endpoint REST y snippet JS opcional para eventos de "lectura"/"visita".
2. Resolución geográfica: servicio de geolocalización por IP con cache y radios de confianza configurables.
3. Agregación: almacenamiento orientado a series temporales (contadores por país/región/ciudad, ventana y buckets).



4. Visualización: mapa interactivo y capa de calor; filtros por rango temporal, país, referrer y path.
5. Privacidad: truncamiento/anonimización de IP, salting, retención limitada y bloqueo de trazas por ruta.

Tecnologías

Backend en Python/Node (API REST), base de datos relacional + clave/valor para cache, job queue para resolución diferida, y Leaflet para cartografía interactiva (tiles libres) con layers de marcadores y calor. Integraciones opcionales con colas (RabbitMQ/Redis) y almacenamiento de eventos en frío.

Recolección de información

No se accede a datos personales adicionales; no se usan cookies de terceros; la IP se trata con técnicas de minimización (truncamiento/anonimización), y se ofrece opt-out por encabezado y parámetro.

Evaluación del software

Pruebas unitarias y de integración; load testing con perfiles de 50–500 rps; verificación de latencia p95 en resolución geográfica; validación visual en navegadores modernos y dispositivos móviles; revisión de accesibilidad (atajos, alt text, contraste).

Ética

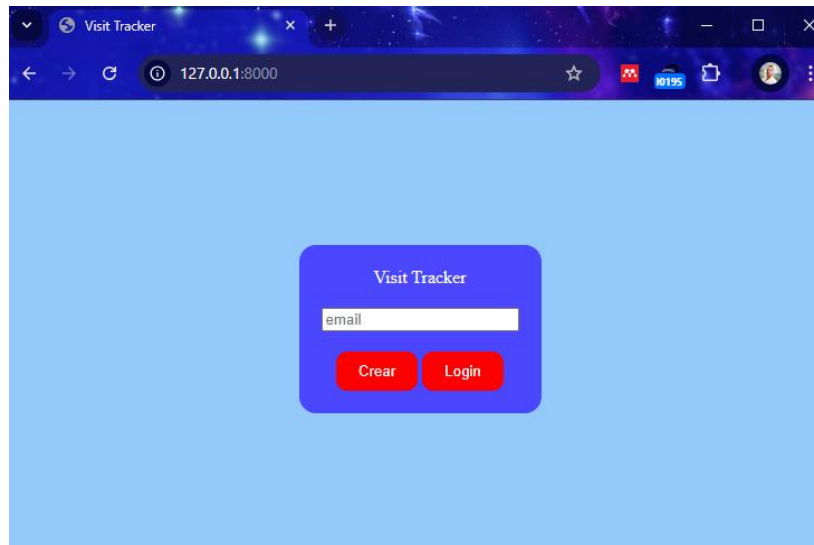
Diseño privacy by default: desactivado el guardado de IP sin anonimización; documentación de las bases legales posibles (interés legítimo/consentimiento según caso) y guías para retención mínima.

RESULTADOS

Interfaz de usuario (UI)

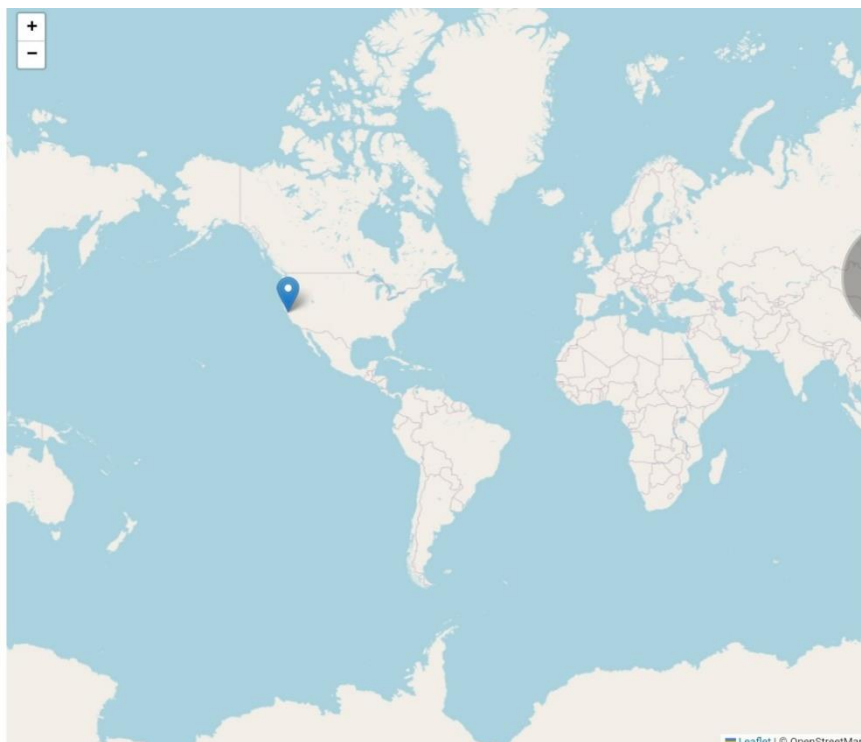
La vista principal cuenta con el formulario de acceso y registro a través del email (Figura 1). Al acceder al usuario se muestra el script que debe ser colocado en el HTML del sitio al que se le desean seguir las estadísticas.

Figura 1. Vista principal de Client Tracker.



En la figura 2 se muestra el mapa resultado, que se observa en el lugar del HTML que se colocó, o como página independiente al acceder a la url de registro con la ruta relativa #id/map.

Figura 2. Vista de mapa de ClientTracker.



Experiencia de usuario(UX)



Atajos de teclado para filtros, tooltips con radio de precisión y confidence, paginación de eventos recientes, y modo oscuro. Documentación integrada con guías de despliegue (README.MD).

Requerimientos

- Hardware/escala: ~1 vCPU y 512–1024 MB RAM para ≤ 50 rps; escala horizontal por workers.
- SO: Linux x86_64/ARM64.
- Dependencias: motor de BD (PostgreSQL/SQLite según escala), servidor de aplicaciones y reverse proxy con TLS.

Pruebas más recientes

- p95 latencia de registro < 40 ms (sin resolución síncrona).
- p95 latencia de resolución geográfica (con cache caliente) ~5–10 ms/consulta.
- Accesibilidad: contraste AA y navegación por teclado en vistas clave.
- Integraciones: webhook hacia tableros de métricas y pipelines ETL.

DISCUSIÓN

La geolocalización por IP ofrece, en general, una alta fiabilidad a nivel de país pero una precisión decreciente al acercarnos a niveles subnacionales; esto se debe a la asignación de bloques de direcciones IP a proveedores o ASNs que sirven amplias áreas geográficas y a la dinámica del ruteo en Internet, por lo que la interpretación de datos a nivel de ciudad requiere precaución metodológica y presentación explícita de incertidumbre en los informes. (1)

Las bases de datos comerciales y abiertas han mejorado su granularidad y sus factores de confianza en los últimos años, pero la variabilidad entre prefijos, la entrada de nuevas redes móviles y el uso creciente de NATs/CG-NAT obliga a desplegar actualizaciones frecuentes y mecanismos de validación cruzada (p. ej., comparar resultados con muestras de clientes voluntarios) para evitar sesgos sistemáticos en los informes de cobertura. (2)

El tratamiento de la dirección IP como dato personal en muchos marcos regulatorios impone obligaciones de transparencia, limitación de la finalidad y



medidas técnicas (anonimización, minimización, retención limitada); así, cualquier despliegue operativo debe acompañarse de registros de tratamiento, evaluaciones de impacto cuando proceda y opciones técnicas de exclusión para usuarios, a fin de reducir riesgos legales y reputacionales. (3,4)

Los cambios en la industria sobre cookies de terceros y las iniciativas de privacidad (como Privacy Sandbox) modifican las palancas disponibles para medir audiencias; por eso, las estrategias de medición basadas exclusivamente en client-side cookies se vuelven frágiles, y conviene diseñar arquitecturas híbridas que combinen señales limpias del servidor con mediciones opcionales en el cliente para mantener continuidad en las series temporales. (5–7)

La instrumentación server-side aporta ventajas concretas: mayor resiliencia ante bloqueadores, control centralizado del procesamiento y posibilidades de aplicar lógica de minimización antes del almacenamiento permanente; sin embargo, exige controles de latencia, mecanismos de deduplicación y transparencia sobre qué eventos se elevan al servidor para evitar sesgos de captura. (8,9)

Las inferencias geográficas sufren distorsiones originadas en la infraestructura (CDN, anycast, túneles VPN, mobile offload), por lo que los análisis deben incorporar métricas de confianza, comparaciones temporales y filtros para detectar picos que correspondan a cambios infraestructurales y no a variaciones reales en la audiencia; sin ese enfoque, se corre el riesgo de tomar decisiones editoriales sobre datos artefactuales. (10)

La elección de la biblioteca cartográfica (Leaflet en este caso) responde a prioridades de ligereza, extensibilidad y compatibilidad con tiles abiertos; optar por una herramienta con amplio ecosistema facilita integraciones, reduce la dependencia de proveedores y permite aplicar capas de datos (por ejemplo, incertidumbre o agrupamientos) sin cargar el cliente con peso innecesario. (11)

En el plano académico y de I+D, hay trabajos recientes que exploran enfoques híbridos —combinando datos de ruteo, reverse DNS, latencias medidas y aprendizaje automático— que muestran potencial para mejorar la precisión en



contextos concretos; dichos enfoques requieren, no obstante, colecciones de entrenamiento representativas y validación ética para evitar amplificar sesgos geográficos. (12)

Finalmente, el debate público y periodístico sobre las promesas y límites de la privacidad en la web ha evidenciado que algunas decisiones regulatorias o de mercado pueden producir efectos secundarios inesperados (p. ej., desplazamiento de técnicas de rastreo hacia huellas más persistentes); por tanto, mantener ClientTracker en modo “mínimos necesarios”, con documentación pública y auditorías técnicas periódicas, es una postura prudente para sostener confianza institucional y cumplir obligaciones emergentes. (13)

CONCLUSIONES

Client Tracker ofrece una alternativa práctica y prudente para mapear la huella geográfica de audiencias académicas mediante una arquitectura privacy-first y modular que comunica incertidumbre, prioriza minimización y retención corta, y combina captura server-side con opciones client-side para mantener continuidad de métricas; su adopción debe acompañarse de actualizaciones frecuentes de las bases de geolocalización, validación cruzada de resultados, políticas claras de gobernanza de datos y revisiones técnicas que mitiguen sesgos derivados de la infraestructura de red, garantizando así una herramienta útil para la toma de decisiones editoriales sin sacrificar cumplimiento ni transparencia.

REFERENCIAS BIBLIOGRÁFICAS

1. MaxMind. Geolocation Accuracy. Support Center[Internet]. 2025 [citado 2025/08/20]. Disponible en: [<https://support.maxmind.com/geolocation-accuracy>]
2. MaxMind. GeoIP release notes — 2024 archive. MaxMind Developer Portal [Internet]. 2024 [citado 2025/08/20]. Disponible en: <https://dev.maxmind.com/geoip/geoip2/geolite2/updates>



3. European Data Protection Board. What is personal data? EDPB SME Guide [Internet]. 2025 [citado 2025/08/20]. Disponible en: [<https://edpb.europa.eu/>]
4. European Data Protection Board. Secure personal data — guidelines. EDPB [Internet]. 2025 [citado 2025/08/20]. Disponible en: <https://edpb.europa.eu/>
5. Google. Third-party cookies | Privacy Sandbox. Privacy Sandbox site [Internet]. 2025 [citado 2025/08/20]. Disponible en: <https://privacysandbox.com/>
6. Segal, J. Britain says Google's online-ad commitments no longer needed. Reuters [Internet]. 2025 [citado 2025/08/20]. Disponible en: <https://www.reuters.com>
7. D'Angelo, A. Inside Google's rocky path towards killing off the web cookie. Business Insider [Internet]. 2024 [citado 2025/08/20]. Disponible en: <https://www.businessinsider.com/>
8. Matomo. Server-side analytics tracking — FAQ. Matomo.org [Internet]. 2024–2025 [citado 2025/08/20]. Disponible en: <https://matomo.org/docs/server-side-tracking>
9. Data Marketing School. Matomo server-side tracking with Google Tag Manager. Data-Marketing-School.com [Internet]. 2024 [citado 2025/08/20]. Disponible en: <https://data-marketing-school.com>
10. Cloudflare. Cloudflare Radar — Year in Review 2023. Cloudflare Blog [Internet]. 2023 Dec [citado 2025/08/20]. Disponible en: <https://blog.cloudflare.com/radar-2023>
11. Leaflet. Leaflet — a JavaScript library for interactive maps. LeafletJS [Internet]. 2025 [citado 2025/08/20]. Disponible en: <https://leafletjs.com/>



12. Zhang X, et al. Accuracy and coverage analysis of IP geolocation databases. Preprint [Internet]. 2023 [citado 2025/08/20]. Disponible en: <http://dx.doi.org/10.1109/BalkanCom58402.2023.10167899>

13. Greenberg A. What Google's U-Turn on Third-Party Cookies Means for Chrome Privacy. WIRED [Internet]. 2024 [citado 2025/08/20]. Disponible en: <https://www.wired.com/story/google-chrome-third-party-cookies-privacy-rollback/>

Los autores certifican la autenticidad de la autoría declarada, así como la originalidad del texto.