



# Semantic Scholar

*Motor de búsqueda académico gratuito impulsado por IA para investigadores, departamentos de I+D+i y analistas de datos. Utiliza procesamiento de lenguaje natural para comprender el contexto semántico de más de 200 millones de artículos científicos, permitiendo identificar citas influyentes, generar resúmenes automáticos TLDR y gestionar bibliotecas personalizadas. Es la herramienta ideal para optimizar revisiones sistemáticas y vigilancia tecnológica en áreas STEM mediante aprendizaje automático.*

[Visitar Sitio Oficial](#) [Preguntar a ChatGPT](#) [Preguntar a Claude](#) [Preguntar a Grok](#)

## Contenido del Dossier

- [Información de la Herramienta](#)
- [Consejos de Implantación](#)
- [Preguntas Frecuentes](#)
- [Contratos y Condiciones](#)

## INFORMACIÓN DE LA HERRAMIENTA

### Qué y para quién es

Semantic Scholar es un motor de búsqueda académico gratuito potenciado por Inteligencia Artificial, desarrollado por el Allen Institute for AI (AI2). A diferencia de los buscadores tradicionales basados en palabras clave, utiliza procesamiento de lenguaje natural (NLP) para comprender el contexto y la semántica de las investigaciones.

Está diseñado específicamente para el ámbito investigador y corporativo que requiere acceso a evidencia científica actualizada: departamentos de I+D+i, analistas de datos, bibliotecarios de investigación, estudiantes de postgrado y desarrolladores que buscan integrar datos bibliométricos en sus propias aplicaciones.

### Principal ventaja profesional

La identificación de "Citas Altamente Influyentes" (Highly Influential Citations). Mediante modelos de aprendizaje automático, la herramienta distingue entre una cita superficial y una donde el trabajo citado ha tenido un impacto real en la metodología o resultados del nuevo estudio, permitiendo filtrar el ruido y centrarse en la literatura realmente relevante.

### Para quién no es

Profesionales que busquen información generalista, noticias, patentes o literatura gris no académica. Tampoco es la herramienta ideal para investigadores en humanidades o artes, ya que su base de datos y sus funciones de IA (como los resúmenes TLDR) están optimizadas principalmente para las áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas).

### Funcionalidades clave

- **TLDR Summaries:** Resúmenes de una sola frase generados por IA que condensan el objetivo y resultados principales de un artículo.
- **Research Feeds:** Canales de recomendación personalizados que aprenden de la biblioteca del usuario para sugerir nuevos artículos publicados.
- **Semantic Reader:** Un lector de PDF enriquecido que muestra definiciones de términos y previsualizaciones de citas sin salir del documento.
- **Gráfico Académico:** Acceso a una red de más de 200 millones de artículos, 79 millones de autores y sus conexiones.
- **Filtros Avanzados:** Segmentación por tipo de publicación, autor, centro de investigación y factor de influencia.

### Precios

- **Versión Gratuita:** El acceso a la plataforma web, la creación de bibliotecas, las alertas y la búsqueda son 100% gratuitos y sin publicidad.
- **API Access:** Existe una modalidad gratuita con límites de tasa (rate-limits) para uso público.
- **API Pro:** Acceso gratuito bajo solicitud para desarrolladores o instituciones que requieran límites de tasa más altos y acceso a datos masivos (bulk data). No existen cuotas mensuales comerciales estándar; es un recurso de ciencia abierta.

### Perfil del usuario

- **Investigadores y Académicos:** Para optimizar revisiones sistemáticas de la literatura.
- **Gestores de Innovación (I+D):** Para vigilar tendencias tecnológicas y descubrimientos científicos que afecten a su industria.
- **Data Scientists:** Para explotar la API y realizar análisis de grafos o minería de textos científicos.
- **Bibliometristas:** Para evaluar el impacto y la evolución de áreas de conocimiento específicas.

### Nivel técnico requerido

- **Uso de la plataforma:** Nivel de usuario básico. La interfaz es intuitiva, similar a otros buscadores académicos.
- **Configuración de la API:** Requiere conocimientos de programación (Python, JavaScript, etc.) y manejo de arquitecturas REST.
- **Competencias necesarias:** Comprensión de la estructura de las publicaciones científicas y capacidad para evaluar la relevancia de los resultados técnicos.

#### Ejemplos de uso profesional

- **Vigilancia Tecnológica:** Un departamento técnico puede configurar "Research Feeds" para recibir automáticamente en su correo las últimas innovaciones en biotecnología o IA sin búsqueda manual.
- **Mapa de Expertos:** Identificar a los autores más influyentes en un campo técnico específico para posibles colaboraciones o contrataciones.
- **Bibliometría Corporativa:** Uso de la API para medir el impacto de las publicaciones producidas por la propia empresa en comparación con la competencia.

#### Uso y distribución

- **Versión web:** Accesible desde cualquier navegador moderno.
- **Versión móvil:** Web adaptativa (no dispone de aplicación nativa dedicada).
- **API:** Interfaz REST para integración programática.
- **Datasets:** Descarga de archivos JSON (S2AG) para análisis offline de grandes volúmenes de datos.

#### Open Source

Aunque la interfaz web no es de código abierto, muchos de los modelos bajo el capó (como SPECTER) y los datasets (S2ORC) se publican bajo licencias abiertas para fomentar la investigación en IA y NLP.

#### Integraciones

- **Facilidad de integración:** De nivel medio (vía API) a alto (uso de datasets).
- **API propia:** REST API documentada para acceder a datos de autores, artículos, citas y embeddings.
- **Ejemplos de integración:** Herramientas como "Connected Papers" o "Litmaps" utilizan la infraestructura de Semantic Scholar para crear visualizaciones de grafos de conocimiento. Se integra con gestores de citas permitiendo exportaciones en BibTeX, MLA, APA y Chicago.

#### Notas finales

##### Información legal, licencias y contratos

- **Licencia API:** El uso de la API requiere la aceptación de un acuerdo de licencia específico que prohíbe el reempaqueado comercial de los datos sin permiso explícito.
- **Propiedad Intelectual:** Los datos de los artículos (S2 Data) se rigen por las licencias originales de sus autores/editoriales (ej. CC BY-NC).
- **Atribución:** Es obligatorio incluir la mención "Powered by Semantic Scholar" en cualquier aplicación o publicación científica que utilice sus datos o API.

#### Para más información:

- Sitio web oficial: <https://www.semanticscholar.org>
- Documentación API: <https://www.semanticscholar.org/product/api>
- Acuerdo de licencia API: <https://www.semanticscholar.org/product/api/license>
- GitHub (Allen Institute for AI): <https://github.com/allenai>

## CONSEJOS DE IMPLANTACIÓN

---

### Aplicación profesional

Semantic Scholar se aplica principalmente en sectores vinculados a la innovación y el desarrollo científico. Es ideal para departamentos de I+D+i en farmacéuticas, biotecnología y tecnología profunda (Deep Tech) que necesitan monitorizar el estado del arte sin el sesgo de buscadores comerciales. Las empresas con presupuesto cero para herramientas de vigilancia tecnológica encuentran aquí un recurso de alta fidelidad. Los puntos clave de aplicación residen en la reducción del tiempo de lectura mediante resúmenes automáticos y la detección de tendencias emergentes a través de su motor de recomendaciones.

### Funcionalidades clave

- [TLDR Summaries](#): Generación automática de resúmenes de una oración que destacan el objetivo y los hallazgos principales, eliminando la necesidad de leer resúmenes extensos en la fase de cribado.
- [Research Feeds](#): Canales de descubrimiento que emplean aprendizaje reforzado para sugerir literatura basada en la biblioteca personal del usuario.
- [Semantic Reader](#): Lector inteligente que resalta menciones y permite ver referencias y definiciones en ventanas emergentes sobre el propio PDF.
- [Highly Influential Citations](#): Algoritmo que identifica si una cita es meramente formal o si el artículo citado es fundamental para la metodología del nuevo estudio.
- [Open Research API](#): Acceso programático para extraer metadatos, grafos de citas y embeddings de más de 200 millones de publicaciones.

### Perfiles necesarios

- Perfiles técnicos: Desarrolladores de software o Data Scientists si se desea explotar la API para integrar datos científicos en sistemas internos de gestión del conocimiento.
- Especialistas sectoriales: Documentalistas, gestores de patentes o investigadores de campo (STEM) para la configuración de alertas y selección de literatura crítica.

### Retorno de la inversión (ROI)

- Tiempos: Se estima una reducción del 30% al 50% en el tiempo dedicado a la revisión bibliográfica inicial gracias a los resúmenes TLDR y al filtrado de citas influyentes.
- Medición de KPIs: Volumen de artículos relevantes encontrados por hora de búsqueda, número de nuevas patentes o proyectos de innovación fundamentados en literatura descubierta y ahorro en costes de licencias de bases de datos científicas comerciales de pago.

### Otros

- Integraciones externas: Semantic Scholar es el motor que alimenta herramientas populares de visualización de datos como Connected Papers y Zotero (vía plugins), lo que permite una integración fluida en el flujo de trabajo de escritura científica ya existente.
- Limitación disciplinar: La efectividad de los modelos de IA es significativamente mayor en Ciencias de la Computación, Biomedicina y Física, siendo menos precisa en Ciencias Sociales y colectando menos datos en Humanidades.

## PREGUNTAS FRECUENTES

---

### ¿Qué es Semantic Scholar y en qué se diferencia de buscadores como Google Scholar?

Semantic Scholar es un motor de búsqueda académico gratuito desarrollado por el Allen Institute for AI (AI2). A diferencia de los motores basados en palabras clave, utiliza Inteligencia Artificial y Procesamiento de Lenguaje Natural (NLP) para comprender el contexto semántico de las investigaciones, permitiendo resultados más precisos y herramientas de análisis avanzado como el filtrado de citas influyentes.

### ¿Qué son las 'Citas Altamente Influyentes' y para qué sirven profesionalmente?

Es una métrica propietaria basada en aprendizaje automático que identifica cuándo un artículo citado ha sido fundamental para la metodología o los resultados del estudio que lo cita. Para un profesional, esto permite filtrar el ruido bibliográfico y priorizar la lectura de los trabajos que realmente han impactado en el avance de una línea de investigación específica.

### ¿Cuáles son los costes asociados y existe una versión gratuita?

El acceso a la plataforma web, la creación de bibliotecas y las herramientas de búsqueda son totalmente gratuitos y sin publicidad. Para el uso técnico, la API ofrece una modalidad gratuita con límites de tasa y una modalidad Pro, también gratuita, sujeta a solicitud para instituciones o investigadores que requieran acceso a datos masivos (bulk data).

### ¿Es una herramienta de código abierto (Open Source)?

La plataforma web no es de código abierto. Sin embargo, el Allen Institute for AI publica bajo licencias abiertas muchos de sus modelos de IA subyacentes, como SPECTER, y sus conjuntos de datos académicos (como S2ORC y S2AG) para fomentar la investigación en minería de textos y ciencia abierta.

### ¿Cómo aborda Semantic Scholar la privacidad y la normativa de datos?

Como proyecto del Allen Institute for AI (una organización sin ánimo de lucro), la plataforma se centra en la difusión de conocimiento científico. El uso de la API requiere la aceptación de términos de licencia específicos que protegen la propiedad intelectual de los autores originales y prohíben la redistribución comercial no autorizada de los datos bibliográficos.

### ¿Qué funcionalidades de IA facilitan la lectura rápida de artículos?

Destacan los 'TLDR Summaries', que son resúmenes de una sola frase generados por IA que condensan el objetivo y hallazgos principales, y el 'Semantic Reader', un lector de PDF que ofrece definiciones contextuales y previsualizaciones de citas integradas para evitar interrupciones durante la investigación.

### ¿Es adecuado para todas las disciplinas académicas?

No es ideal para todas. Aunque cubre una vasta red de 200 millones de artículos, sus modelos de IA y bases de datos están optimizados principalmente para áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), pudiendo ser menos eficaz en disciplinas de Humanidades o Artes.

### ¿Ofrece integración con otras herramientas profesionales?

Sí, permite la exportación de citas a gestores bibliográficos en formatos BibTeX, MLA, APA y Chicago. Además, mediante su API REST, se integra con herramientas de visualización de grafos de conocimiento como Connected Papers y Litmaps, y permite a desarrolladores extraer metadatos de autores, artículos y embeddings.

### ¿Se puede utilizar con fines comerciales?

El uso de los datos está sujeto a la licencia de Semantic Scholar, que generalmente permite la investigación y el desarrollo de herramientas, pero prohíbe el reempaquetado comercial de la base de datos sin permiso explícito. Cualquier aplicación que utilice su tecnología debe incluir obligatoriamente la atribución 'Powered by Semantic Scholar'.

### ¿Requiere instalación de software?

No, es una plataforma basada en la nube accesible desde cualquier navegador moderno. No dispone de una aplicación móvil nativa, pero su interfaz es web adaptativa, permitiendo su uso en dispositivos móviles a través del navegador.

## CONTRATOS Y CONDICIONES

### Principales recomendaciones

- **Uso institucional:** Se recomienda registrarse con el correo corporativo/institucional para centralizar el acceso a suscripciones y recursos compartidos por la organización.
- **Verificación de resultados:** Al ser una herramienta basada en modelos de IA y LLM (como GPT-3.5 para ciertas funciones), los resúmenes y respuestas pueden contener errores factuales o alucinaciones. Es imperativo contrastar los datos antes de incluirlos en informes técnicos o científicos.
- **Uso de la API:** En integraciones profesionales, asegúrese de no compartir el **API Key** fuera de los usuarios autorizados de la empresa para evitar la suspensión de la cuenta.
- **Atribución obligatoria:** Cualquier uso público de los datos o resultados obtenidos debe incluir la mención "Powered by Semantic Scholar" y citar a "The Semantic Scholar Open Data Platform" en publicaciones científicas.

### Ley de Inteligencia Artificial (AI Act)

- **Clasificación de riesgo:** Según el marco de la UE, Semantic Scholar se clasifica mayoritariamente como IA de **Riesgo Mínimo**, ya que es un motor de búsqueda y herramienta de apoyo a la investigación.
- **Transparencia:** La herramienta cumple con los principios de transparencia al identificar claramente funciones como "TLDR" o "Ask This Paper" como generadas por tecnología IA.
- **Modelos de Propósito General:** Al utilizar modelos externos (como los de OpenAI) para resúmenes y definiciones, la empresa debe ser consciente de que el procesamiento de lenguaje natural está sujeto a las limitaciones de dichos modelos.

### Privacidad y protección de datos

- **Responsabilidades:** El Allen Institute for AI (Ai2) actúa como **Responsable del Tratamiento** (Data Controller) de los datos de los usuarios.
- **Ubicación de los datos:** Los servidores están ubicados en **Estados Unidos**.
- **Transferencia internacional:** El uso de esta herramienta implica una transferencia internacional de datos fuera del Espacio Económico Europeo. La política de privacidad menciona el cumplimiento de estándares razonables, pero no detalla explícitamente la adhesión al Marco de Privacidad de Datos UE-EE. UU. (Data Privacy Framework).
- **Derechos ARCO:** Los usuarios pueden ejercer sus derechos de acceso, rectificación y supresión enviando un correo a [privacy@allenai.org](mailto:privacy@allenai.org). Al eliminar una cuenta, los datos se borran de las bases de datos activas, aunque pueden permanecer en copias de seguridad por motivos legales.

### Propiedad intelectual

- **Propiedad de datos:** El usuario mantiene la propiedad de los datos que introduce (datos de consulta, configuraciones de feed).
- **Propiedad del resultado:** Los datos bibliométricos y metadatos pertenecen a Ai2 y sus licenciantes. Los textos completos de los artículos están sujetos a las licencias originales de sus respectivos autores o editoriales (ej. CC BY-NC).
- **Derechos de autor:** Semantic Scholar no otorga permisos para distribuir o publicar partes de los artículos indexados; el usuario debe contactar directamente con el autor o la editorial.

### Usos y prohibiciones

- **Usos admitidos:** Investigación académica, educación, evaluación de modelos de aprendizaje automático y vigilancia tecnológica profesional.
- **Usos prohibidos:** Reempaquetar o revender la API, eliminar avisos de copyright, realizar scraping abusivo que ignore los límites de tasa (Rate Limits) y utilizar la herramienta para spam, phishing o actividades ilegales.
- **Uso comercial:** La licencia estándar de la API es para fines no comerciales. Para usos comerciales integrados en productos de pago, se requiere solicitar una licencia expandida específica.

### Seguridad y certificaciones

- **Seguridad:** Utiliza cifrado en tránsito y en reposo para sistemas sensibles (como HISE), y medidas técnicas razonables para el resto de sitios.
- **Certificaciones:** El instituto realiza auditorías financieras anuales (Single Audit Reports), pero no se publicita explícitamente una certificación SOC2 o ISO 27001 para el servicio Semantic Scholar.

### Otros

- **Interrupción del servicio:** El servicio se ofrece "tal cual" (as is). Ai2 se reserva el derecho de suspender el acceso a la API sin previo aviso si se detecta un uso abusivo o riesgo de seguridad.

Fuentes consultadas:

- [Acuerdo de licencia de la API](#)
- [Política de privacidad de Ai2](#)
- [Términos de uso y condiciones](#)
- [Directrices de uso responsable](#)
- [Información sobre permisos de contenido](#)

### Para más información y herramientas:

Explora look4.tools para descubrir las mejores soluciones tecnológicas del mercado.

[Inicio](#) [Todas las herramientas](#) [Categorías](#)

Este documento ofrece recomendaciones generadas mediante análisis humano y sistemas de IA automatizados. La información tiene carácter meramente informativo y no constituye asesoramiento legal, profesional ni garantía de resultados. Las marcas, logotipos y nombres comerciales pertenecen a sus respectivos propietarios y se utilizan únicamente con fines identificativos.