

GoPubMed, PubMed y PubMedReMiner: dos herramientas para los análisis métricos y el descubrimiento de conocimientos en PubMed

GoPubMed, PubMed and PubMedReMiner: two useful tools for metric analysis and knowledge disclosing in PubMed

Rubén Cañedo Andalia^I; Sonia Santana Arroyo^{II}; Javier Santovenia Díaz^{III}

^ILicenciado en Información Científico-Técnica y Bibliotecología. Departamento Fuentes y Servicios de Información. Centro Nacional de Información de Ciencias Médicas-Infomed.La Habana.Cuba.

^{II}Licenciada en Información Científico-Técnica y Bibliotecología. Biblioteca Médica Nacional. Centro Nacional de información de Ciencias Médicas-Infomed.La Habana.Cuba.

^{III}Licenciado en Información Científico-Técnica y Bibliotecología. División Autopartes. Unecamoto.La Habana. Cuba.

Durante casi dos décadas se han realizado muchos intentos por desarrollar herramientas para la realización de análisis métricos de las bases de datos *Medline* y *PubMed*. Algunos han hecho esos análisis incluso en forma manual o semiautomática. El objetivo de este trabajo es presentar dos herramientas basadas en Web, gratuitas, y que constituyen dos joyas para la minería de datos en *PubMed*: *GoPubMed* y *PubMed PubReMiner*

GoPubMed

GoPubMed (<http://www.gopubmed.com/>) es un buscador para la Web 2.0, diseñado por la Universidad Técnica de Dresde, Alemania y la empresa Transinsight GmbH,¹ dedicada al desarrollo de tecnologías de búsqueda inteligentes (semánticas) para las ciencias de la vida ([figura 1](#)). Este buscador posibilita una rápida y fructífera exploración bibliográfica de la base de datos *PubMed*, que hoy día acumula 18 millones de referencias; así como un análisis métrico inmediato de los resultados obtenidos en las búsquedas realizadas.

El sistema, basado en una ontología genética -*Gene Ontology* (GO) (<http://www.geneontology.org/>), es en esencia un vocabulario controlado concebido para la búsqueda de productos genéticos, desarrollados por *Gene Ontology Consortium*- y el MeSH, el tesoro de la Biblioteca Médica Nacional de los

Estados Unidos de América, que se utiliza para la indización y recuperación de la información disponible en *PubMed*, su base de datos insignia, organiza por categorías o clases temáticas las referencias recuperadas y, sobre la base de esa clasificación, presenta los resultados a sus usuarios.

Para realizar una búsqueda mediante *GoPubMed*, se debe teclear una estrategia - por principio cualquier estrategia procesable por *PubMed* lo es también por *GoPubMed*- en la ventana que ofrece el sistema como si se trabajara directamente con la interfaz de *PubMed* (*Entrez*).

Una vez realizada la búsqueda, si por ejemplo se ha tecleado el término "leprosy", el sistema presenta sus resultados en cuatro acápites principales: *What*, *Who*, *Where* and *When*. En el primero aparecen a la izquierda los resultados de la búsqueda, agrupados temáticamente en clases, como se había referido anteriormente, según las jerarquías controladas de la GO y el MeSH. Cada una de estas clases se puede volver a explorar tantas veces como sea necesario hasta alcanzar la precisión deseada en los resultados de la búsqueda ([figura 2](#)).

GoPubMed entrega por omisión hasta un total de 1 000 referencias; distingue al autor más productivo del conjunto recuperado; informa sobre la existencia de resumen para cada referencia; destaca la palabra buscada en el texto; presenta la idea central con la que concluye el artículo; posibilita la eliminación de referencias inútiles y facilita varias claves para realizar nuevas exploraciones, entre ellas el autor y el título de la revista.

En el segundo acápite, *Who*, es posible precisar los autores más productivos; así como observar las redes sociales de esos especialistas ([figura 3](#)). Por primera vez un software basado en Web extrae de manera totalmente automática y libre las redes sociales de colaboración científica a partir de una gran base de datos como *PubMed*.² En el tercero, *Where*, es posible hallar las revistas más productivas. En el cuarto y último se puede apreciar una distribución de las publicaciones según años de edición.

Si se presiona en la columna principal la opción *Show statistics for this 1,000 articles*, el sistema entregará un paquete de tablas y gráficos con información sobre los países y las ciudades más productivas para la temática ([figura 4](#)), los títulos de revistas más productivos en la temática y la distribución de la producción según años ([figura 5](#)), el comportamiento del interés en la temática en el período consultado ([figura 6](#)) y la distribución de la producción sobre la temática consultada en un mapamundi ([figura 7](#)).

Al seleccionar *Hot Topic in Go & MeSH*, se puede conocer de inmediato el tema más frecuente, en este caso *Leprosy* como enfermedad. El sistema ofrece entonces una definición de esta afección y, a continuación, presenta una relación de los autores, las revistas, las ciudades y los países más productivos en la temática, además de su comportamiento durante los últimos 20 años ([figuras 8,9,10](#)).

PubMed PubReMiner

PubMed PubReMiner (<http://bioinfo.amc.uva.nl/human-genetics/pubreminer/>), diseñado por Jan Koster, del *Department of Human Genetics del Academic Medical Center de Amsterdam*,³ es un asistente para la búsqueda en *PubMed* ([figura 11](#)), que permite iniciar la exploración de dicha base de datos a partir de un palabra cualquiera ([figuras 12](#) y [13](#)). Combinada sucesivamente con otras palabras y claves para la búsqueda, como son los nombres de los autores, los títulos de las revistas, los años, los países, las sustancias o los aspectos más tratados, posibilita

finalmente alcanzar un resultado que se ajusta mejor a los intereses bibliográficos de quienes utilizan el sistema ([figura 14](#)). Al finalizar la búsqueda con el asistente, es posible dirigirse a *PubMed* mediante un clic sobre el botón *Goto PubMed with query*. Los resultados del proceso estadístico se pueden guardar en un fichero *.txt*.

PubMed PubReMiner genera con rapidez estadísticas métricas esenciales asociadas a los resultados de una solicitud de búsqueda específica realizada para *PubMed*. Dichas estadísticas comprenden fundamentalmente las revistas, los autores y países más productivos en el conjunto estudiado; la distribución de la producción según años y los términos o palabras (palabras clave, términos MeSH y nombres de sustancias) utilizadas con mayor frecuencia en los títulos, campos de descriptor y de sustancias; así como en los resúmenes de los trabajos recuperados.

La introducción de una estrategia tan simple como *cuba [AD]* en la ventana de búsqueda de *PubMed PubReMiner*, posibilita en unos minutos determinar en *PubMed* el total de la producción de contribuciones, en las cuales su primer autor pertenece a una institución cubana. Es importante que esto no se confunda con la producción total de autores cubanos en esta base de datos; así como realizar una breve panorámica de dicha producción. Por ejemplo, el total de registros de contribuciones de primeros autores cubanos en *PubMed* alcanza la cifra de 3 668 artículos (9 de agosto de 2008). Desde la década de 1990 hasta 2004 hubo tendencia al incremento de la presencia de la producción de la isla en la base de datos. Sin embargo, a partir de ese momento se observa un leve descenso de esta. Un total de 5 autores han publicado 50 o más trabajos. Ellos son: *R. Más, MG. Guzmán, R. Pérez, M. Noa y M Álvarez*. La *Revista de Neurología*, la *Revista Cubana de Medicina Tropical* y los *Archivos Españoles de Urología* acumularon cada una más de 100 contribuciones. Se aprecia un predominio de los estudios en humanos sobre los realizados con animales y que los anticuerpos monoclonales y las proteínas recombinantes fueron las sustancias más estudiadas ([figura 15](#)).

Tanto *GoPubMed* como *PubMed PubReMiner* facilitan sustancialmente responder preguntas como: ¿cuáles son los autores más productivos en un tema específico?, ¿cuáles revistas publican con mayor frecuencia contribuciones en el tema consultado?, ¿cuáles revistas son las más apropiadas para enviar una contribución por su afinidad con el tema?, ¿cuáles subtemáticas se tratan con mayor frecuencia?, ¿cuál es la tendencia a la publicación de contribuciones sobre el tema según años?

Asimismo posibilitan la respuesta a interrogantes más específicas como: ¿cuáles enfermedades se relacionan con determinado microorganismo?, ¿cuáles son los centros y autores líderes en una técnica en un país determinado?, ¿cuáles temas trabaja un autor o grupo particular?, ¿con cuáles enfermedades se relaciona determinado gen o molécula?^{1,2}

Ambas son formidables herramientas para descubrir información (minería de datos) en una gran base de datos, que es imposible obtener a partir de la simple consulta de *PubMed*. Su importancia para los bibliotecarios y médicos es incalculable y significa un ahorro de miles de horas de trabajo y otro tanto en la inversión de recursos y tiempo en investigaciones métricas para situaciones particulares. Posibilita entonces la realización masiva de estudios métricos "limitados", cuya ejecución, aún con esa condición de "limitados", constituye hasta el momento una clase de análisis privativo generalmente del quehacer de especialistas en información. Con frecuencia estos se desarrollan en forma semiautomática y fuera del Web, a partir de la descarga y posterior procesamiento de los datos con programas diversos, cada uno con fines particulares con vistas a complementar sus resultados.

La realización de estudios métricos amplios y profundos -en contraposición a estos que se han denominado aquí "limitados"- puede requerir del concurso de diversos especialistas y de múltiples estrategias de búsqueda; de su comparación y complementación con vistas a elevar su eficacia; de la verificación y normalización manual de extensos volúmenes de datos; así como de análisis cualitativos con vistas a explicar el comportamiento del flujo analizado. Estos permanecerán sin duda durante algún tiempo más en manos de grupos especializados en la realización de esta clase de análisis.

Sin embargo, gracias a estos sistemas antes referidos, el usuario de *PubMed* puede ahora valorar, desde una perspectiva métrica ajustada a sus intereses personales, el panorama bibliográfico sobre un tema determinado y orientarse con facilidad en su campo con respecto a los autores más productivos, las revistas donde con mayor frecuencia aparecen contribuciones sobre su tema, los países e instituciones que marchan a la cabeza de las investigaciones, etcétera.

Dominar totalmente estas herramientas llevará cierto tiempo pero, una vez logrado esto, se podrá explotar significativamente mejor la información contenida en *PubMed* y se estará en condiciones de descubrir cada cual por sí mismo nuevos conocimientos a partir de la información almacenada en esta gran base de datos. Para quienes comienzan, es un paso esencial en este sentido dominar los recursos para la elaboración de estrategias de búsqueda que ofrece *PubMed*. Una buena guía para ello puede ser *Buscar en Medline con Pubmed: guía de uso en español*.⁴ Los resultados finales que cada uno obtenga con su explotación, dependerán en gran medida de su iniciativa creadora y del empeño por conocer el total de posibilidades que ofrece cada herramienta.

A pesar de todo lo expresado, una última advertencia es siempre saludable: estos sistemas son perfectibles. Una primera apreciación revela que pueden ser muy útiles, tanto para usuarios como para bibliotecarios. *GoPubMed* es muy rápido en su análisis y sus resultados se presentan de una manera más comprensible a primera vista, pero su alcance es más limitado. Si bien *PubMed PubReMiner* es algo más lento, puede analizar hasta 10 000 registros. Entonces serán sólo el tiempo y los estudios, incluso las comparaciones sucesivas que se realicen, los que dirán la última palabra respecto a la utilidad, el alcance y la confiabilidad de sus resultados. Pero esto ¡es tarea de todos!

REFERENCIAS BIBLIOGRÁFICAS

1. Kovic I. PubReMiner. Disponible en: <http://ivor-kovic.com/blog/?p=89> [Consultado: 29 de julio de 2008].
2. Transinsight. Red social de expertos biomédicos integrada en *GoPubMed* de Transinsight. Disponible en: http://www.transinsight.com/media/pdf/Social_Web_for_Bio_PR_ES.pdf [Consultado: 30 de julio de 2008].
3. Kovic I. GoPubMed. Disponible en: <http://ivor-kovic.com/blog/?p=128> [Consultado: 1 de agosto de 2008].
4. Fistera.com. Buscar en Medline con Pubmed: guía de uso en español. Disponible en: http://www.fistera.com/recursos_web/no_explor/pubmed.asp [Consultado: 1 de agosto de 2008].

Recibido: 11 de agosto de 2008.
Aprobado: 16 de agosto de 2008.

Lic. Rubén Cañedo Andalia. Departamento Fuentes y Servicios de Información.
Centro Nacional de Información de Ciencias Médicas-Infomed. Calle 27 No. 110 e/
N y M, El Vedado. Plaza de la Revolución. Ciudad de La Habana. Cuba. Correo
electrónico: ruben@infomed.sld.cu

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS¹

ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN; BIBLIOMETRÍA; BASES
DE DATOS BIBLIOGRÁFICAS.

INFORMATION STORAGE AND RETRIEVAL; BIBLIOMETRICS; DATABASES,
BIBLIOGRAPHIC.

Según DeCI²

ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN; PROCESAMIENTO DE
LA INFORMACIÓN; INFORMETRÍA; MINERÍA DE DATOS; BASES DE DATOS
BIBLIOGRÁFICAS; ARTICULOS CIENTIFICOS/análisis; MEDLINE.
INFORMATION STORAGE AND RETRIEVAL; INFORMATION PROCESSING;
INFORMETRICS; DATA MINING; DATABASES, BIBLIOGRAPHIC; SCIENTIFIC
ARTICLES/analysis; MEDLINE.

¹BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.

Disponible en: <http://decs.bvs.br/E/homepagee.htm>

²Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la
Información. Descriptores en Ciencias de la Información (DeCI). Disponible en:
<http://cis.sld.cu/E/tesauro.pdf>

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos
de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual
2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar
los resultados del trabajo en la práctica, así como todos sus derivados, sin
propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente
el autor o los autores y su fuente original.

Cita (Vancouver): Cañedo Andalia R, Santana Arroyo S, Santovenia Díaz J. GoPubMed y PubMedReMiner: dos herramientas para los análisis métricos y el descubrimiento de conocimientos en PubMed. *Acimed* 2008; 18(3). Disponible en: Dirección electrónica de la contribución. [Consultado: día/mes/año].

GoPubMed - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Ir Vínculos Google G Settings

Dirección http://www.gopubmed.com/

what

Top categories

- Diseases [948]
 - Leprosy [825]
 - Leprosy, Lepromatous [149]
 - Tuberculosis [149]
 - Leprosy, Borderline [149]
 - Leprosy, Tuberculoi [149]
 - more
- Organisms [620]
 - Mycobacterium lepra [282]
 - Mycobacterium [282]
 - more
- Named Groups [609]
 - Patients [463]
 - more
- Chemicals and Drugs [56]
 - Leprostatic Agents [1]
 - more
- Biological Sciences [769]
 - Prevalence [106]
 - more
- Techniques and Equipme [282]
- Health Care [893]
- biological_process [360]
- Anatomy [387]
- Natural Sciences [499]
 - cellular_component [135]
 - molecular_function [73]
- Technology, Industry, Ac [282]
- Psychiatry and Psycholo [282]
- Unclassified [20]

Hot topics in GO & MeSH

- Leprosy
- Find categories ...
- Find related categories

1,000 articles

PubMed has found 20,045 citations for the query **leprosy**. The 1,000 latest documents were used by GoPubMed.

Show statistics for these 1,000 articles.

Katoch, V M – Agra, India [Correct this profile](#) [Learn more](#)

Show all 144 articles of this author.

URL to this profile

Internationally leading author in [Mycobacterium lepra](#), [Leprosy](#), [Leprostatic Agents](#), [Leprosy](#), [Lepromatous](#), [Mycobacterium](#), [Leprosy](#), [Borderline](#), [Clofazimine](#), [Dapsone](#), [Rifampin](#)

[1987 - 2007] Department of Microbiology, Central JALMA Institute for Leprosy, Taj Ganj, Agra.

To see other authors, see the **who** section on the left side.

14: **Leprosy** occurring as immune reconstitution syndrome. PMID: 18639911 [Related Articles](#)

Deps PD et al., *Trans R Soc Trop Med Hyg*, 2008

Leprosy has been reported as an example of an IRIS, and it is important that this **syndrome** should be recognized in **leprosy**-endemic areas.

6: **Leprosy**-specific oral lesions: a report of three cases. PMID: 18667979 [Related Articles](#)

Motta AC et al., *Med Oral Patol Oral Cir Bucal*, 13 (8): E479-82, 2008

The clinical spectrum of **leprosy** ranges from the tuberculoid form (TT) to the disseminative and progressive lepromatous form (LL).

Listo

Inicio Ivor Kovic, Phot... GoPubMed - Mic... PubMed PubReM... GoPubMed - Mi... Novedad de Sonia PARA ACIMED - ... 16:47

GoPubMed - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Ir Vinculos Google Settings

Dirección http://www.gopubmed.com/

who

- Top authors for query
 - Katoch V [29]
 - Sarno E [24]
 - Sherma V [21]
 - Lockwood D [21]
 - Ishii N [19]
 - more
- Find specific author

where

- Geo
 - Journals
 - Top journals for query
 - Lepr Rev [130]
 - Indian J Dermatol Ven [44]
 - Indian J Lepr [44]
 - Nihon Hansenbyo Ge [22]
 - Int J Dermatol [22]
 - more
 - High impact journals only
 - Find specific journal ...
 - Reviews only
 - Internet & more
 - Advanced

when

- Top years for query
 - 2007 [370]
 - 2006 [348]
 - 2008 [174]
 - 2005 [108]
 - Publication date

leprosy find it! go pubmed

1,000 articles

PubMed has found 20,045 citations for the query **leprosy**. The 1,000 latest documents were used by GoPubMed.

Show statistics for these 1,000 articles.

Katoch, V M – Agra, India [Correct this profile](#) [Learn more](#)

Show all 144 articles of this author.

URL to this profile

Internationally leading author in [Mycobacterium leprae](#), [Leprosy](#), [Leprostatic Agents](#), [Leprosy](#), [Lepromatous Mycobacterium Leprosy](#), [Borderline Clofazimine Dapsone Rifampin](#)

[1987 - 2007] Department of Microbiology, Central JALMA Institute for Leprosy, Taj Ganj, Agra.

To see other authors, see the **who** section on the left side.

14: **Leprosy** occurring as immune reconstitution syndrome. [PMID: 18639911](#) [Related Articles](#)

Deps PD et al., *Trans R Soc Trop Med Hyg*, 2008

Leprosy has been reported as an example of an IRIS, and it is important that this **syndrome** should be recognized in **leprosy**-endemic areas.

6: **Leprosy**-specific oral lesions: a report of three cases. [PMID: 18667979](#) [Related Articles](#)

Motta AC et al., *Med Oral Patol Oral Cir Bucal*, 13 (8): E479-82, 2008

The clinical spectrum of **leprosy** ranges from the tuberculoid form (TT) to the disseminative and progressive lepromatous form (LL).

Listo

Inicio Ivor Kovic, Phot... GoPubMed - Micr... PubMed PubReM... GoPubMed - Mi... Novedad de Sonia PARA ACIMED - ... 17:10

GoPubMed - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Dirección <http://www.gopubmed.com/>

leprosy [find it!](#)

1,000 articles

PubMed has found 20,045 citations for the query **leprosy**. The 1,000 latest documents were used by GoPubMed.

[Show statistics for these 1,000 articles.](#)

Katoch, V M – Agra, India [Correct this profile](#) [Learn more](#)

[Show all 144 articles of this author.](#)

[URL to this profile](#)

Internationally leading author in [Mycobacterium leprae](#), [Leprosy](#), [Leprostatic Agents](#), [Leprosy](#), [Lepromatous Mycobacterium Leprosy](#), [Borderline Clofazimine Dapsone Rifampin](#)

[1987 - 2007] Department of Microbiology, Central JALMA Institute for Leprosy, Taj Ganj, Agra.

To see other authors, see the **who** section on the left side.

14: Leprosy occurring as immune reconstitution syndrome. [PMID: 18639911](#) [Related Articles](#)

[Deps PD et al.](#), [Trans R Soc Trop Med Hyg.](#), 2008

Leprosy has been reported as an example of an IRIS, and it is important that this **syndrome** should be recognized in **leprosy**-endemic areas.

6: Leprosy-specific oral lesions: a report of three cases. [PMID: 18667979](#) [Related Articles](#)

[Motta AC et al.](#), [Med Oral Patol Oral Cir Bucal.](#) 13 (8): E479-82, 2008

The clinical spectrum of **leprosy** ranges from the tuberculoid form (TT) to the disseminative and progressive lepromatous form (LL).

who

- Top authors for query
- Katoch V [29]
- Sarno E [24]
- Sherma V [21]
- Lockwood D [21]
- Ishii N [19]
- more
- Find specific author

where

- Geo
- Journals
- Top journals for query
- Lepr Rev [130]
- Indian J Dermatol Venereol Leprol [44]
- Indian J Lepr [44]
- Nihon Hansenbyo Gakkaishi [22]
- Int J Dermatol [22]
- more
- High impact journals only
- Find specific journal ...
- Reviews only
- Internet & more
- Advanced

when

- Top years for query
- 2007 [370]
- 2006 [348]
- 2008 [174]
- 2005 [108]
- Publication date

Inicio Ivor Kovic, Phot... GoPubMed - Micr... PubMed PubReM... GoPubMed - Mi... Novedad de Sonia PARA ACIMED - ... 17:10

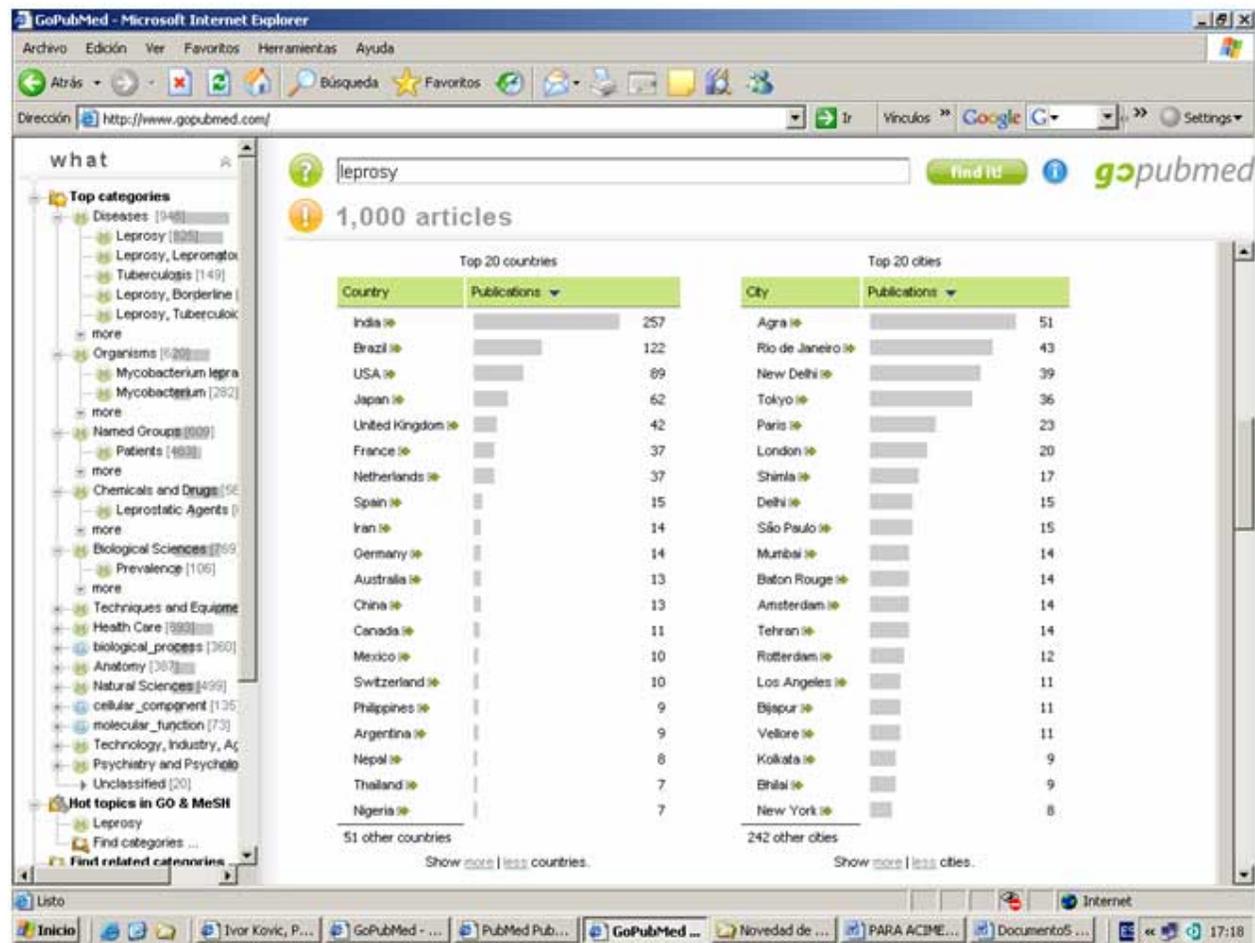


Fig. 4. Países y ciudades más productivas en la temática consultada.

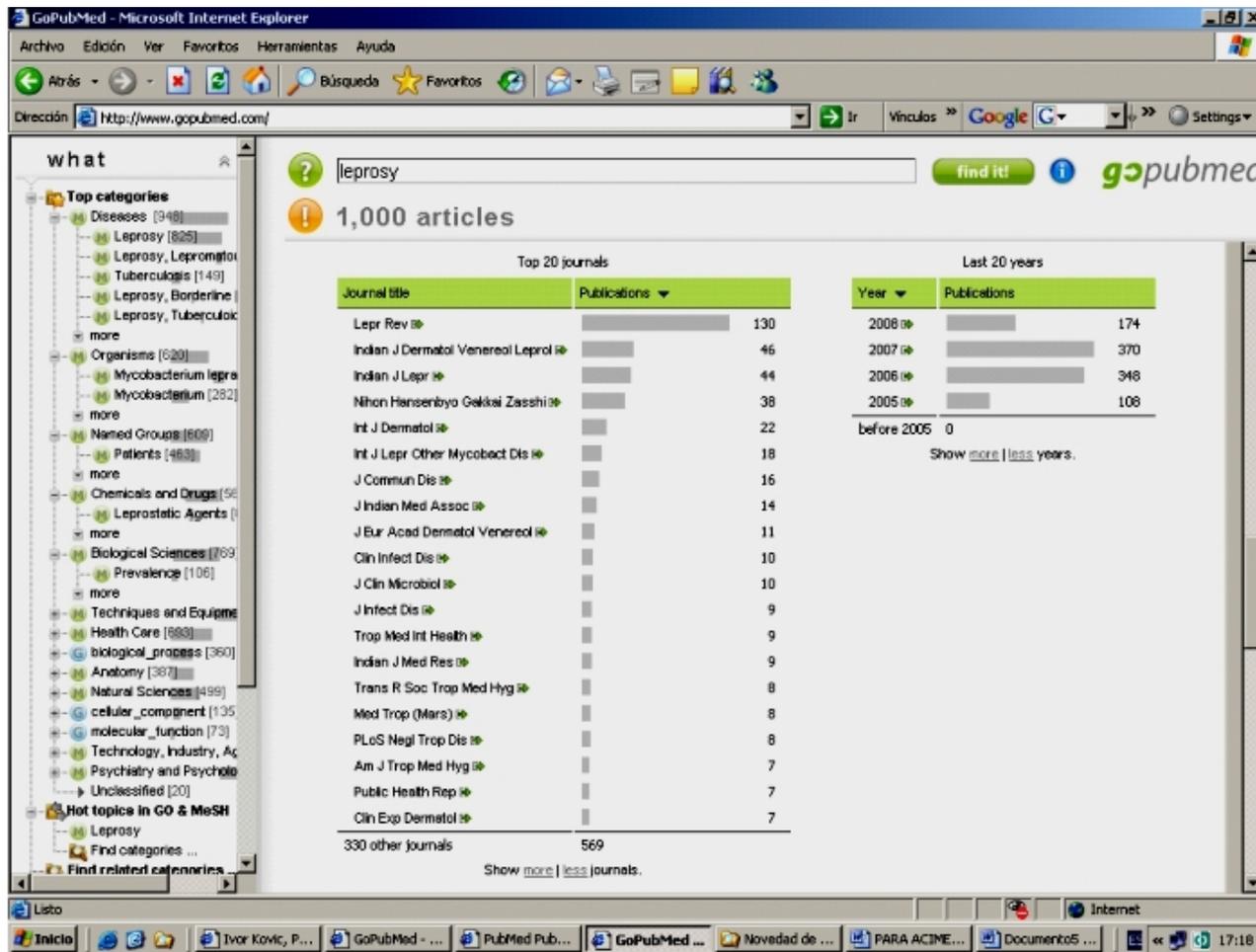


Fig.5. Títulos de revistas más productivos en la temática y distribución de la producción según años.

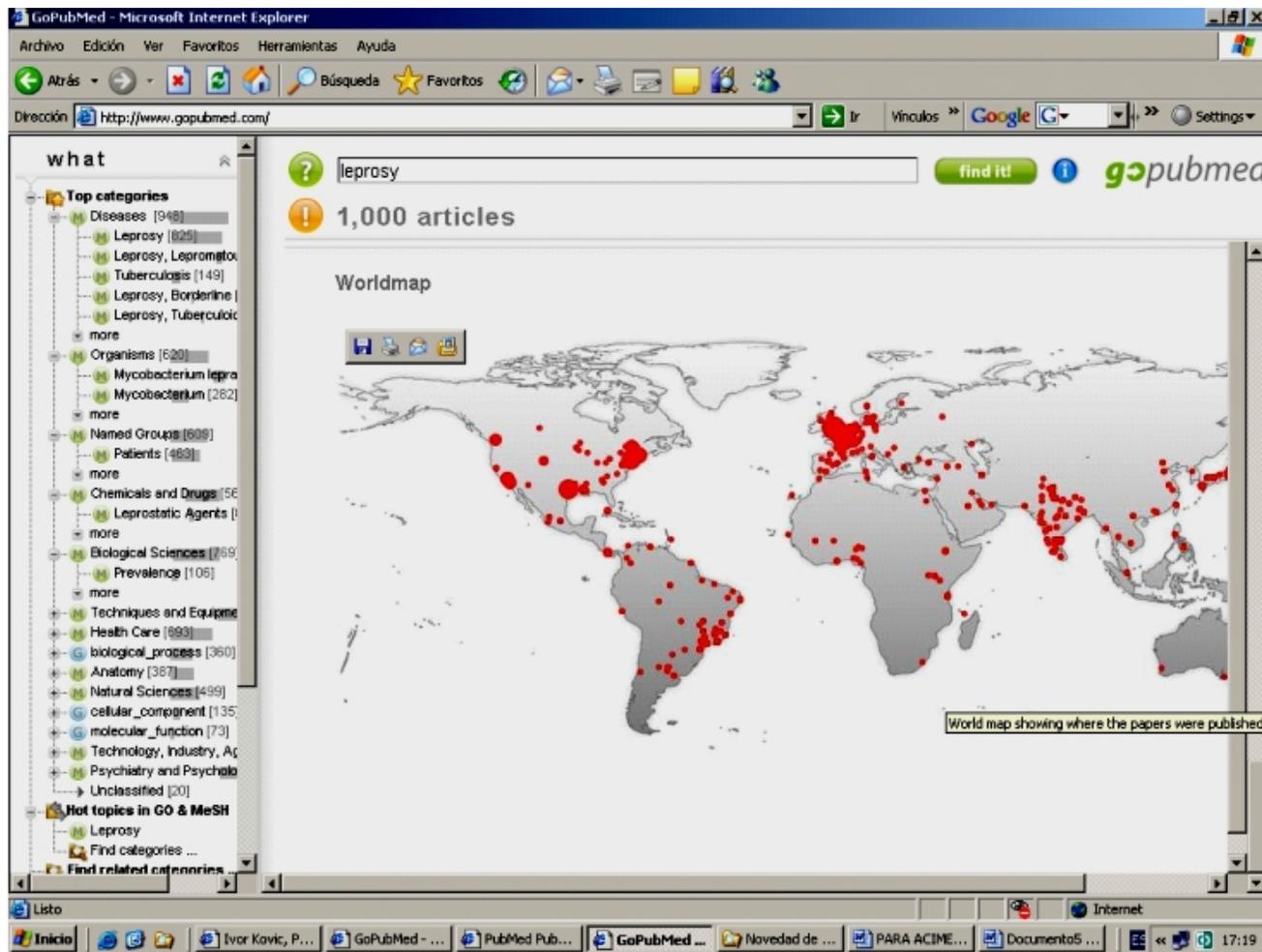


Fig.7. Distribución de la producción sobre la temática consultada en un mapamundi.

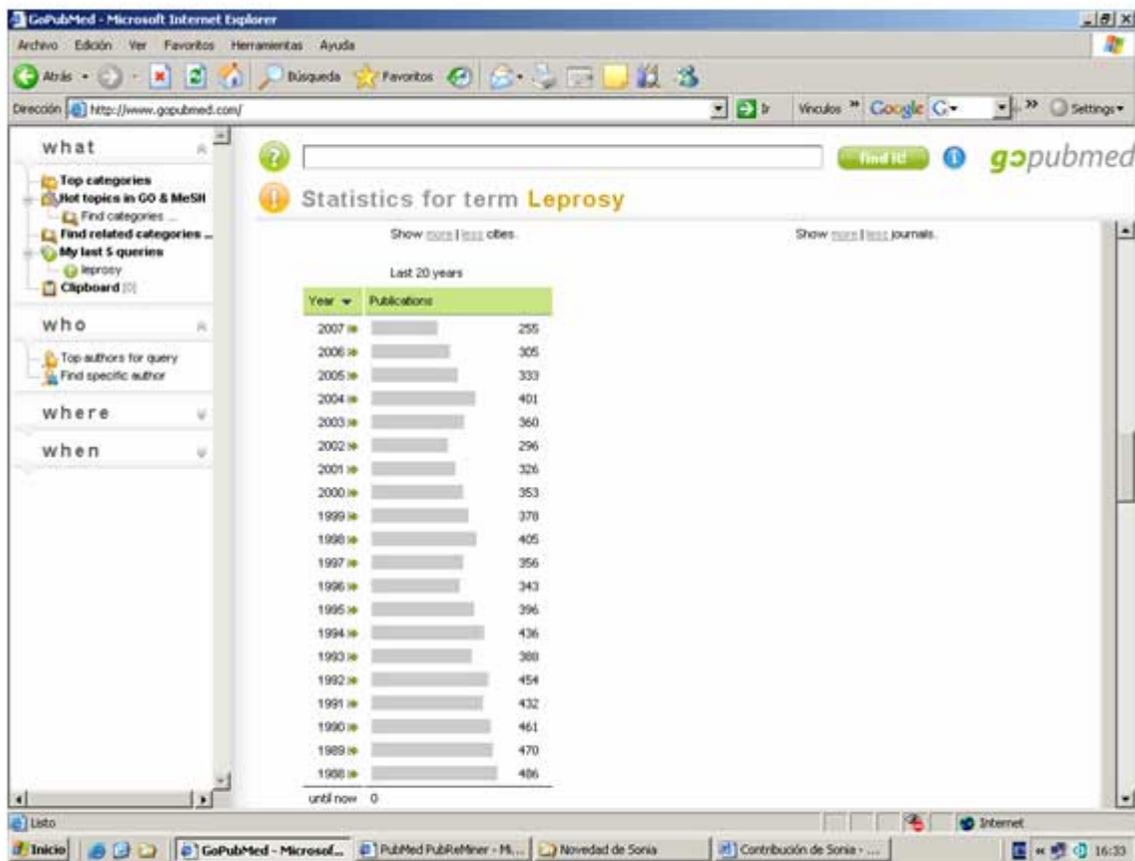


Fig. 8. Comportamiento de la temática más frecuente durante los últimos 20 años.

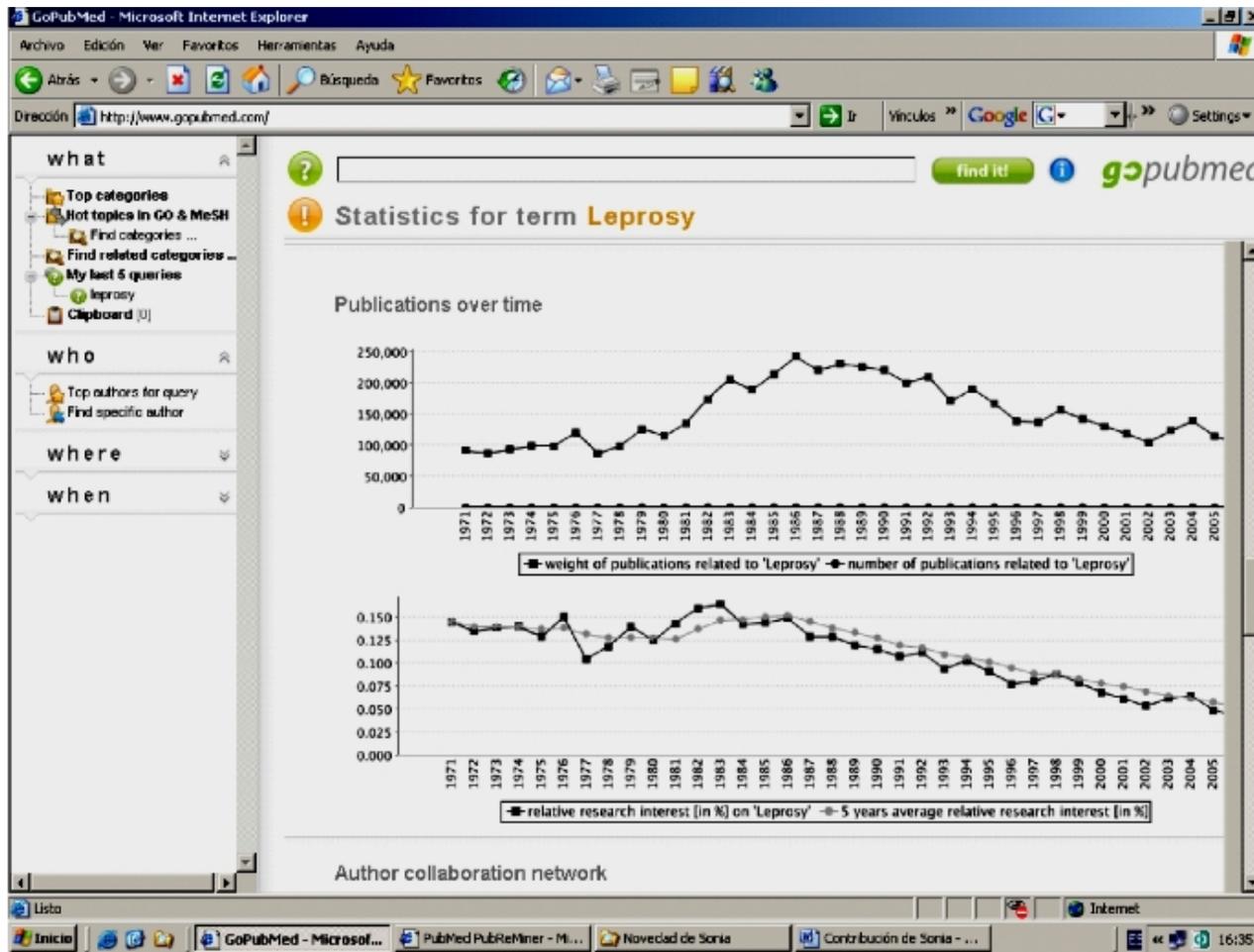


Fig. 9. Gráfico de comportamiento de la temática *Leprosy* como enfermedad.

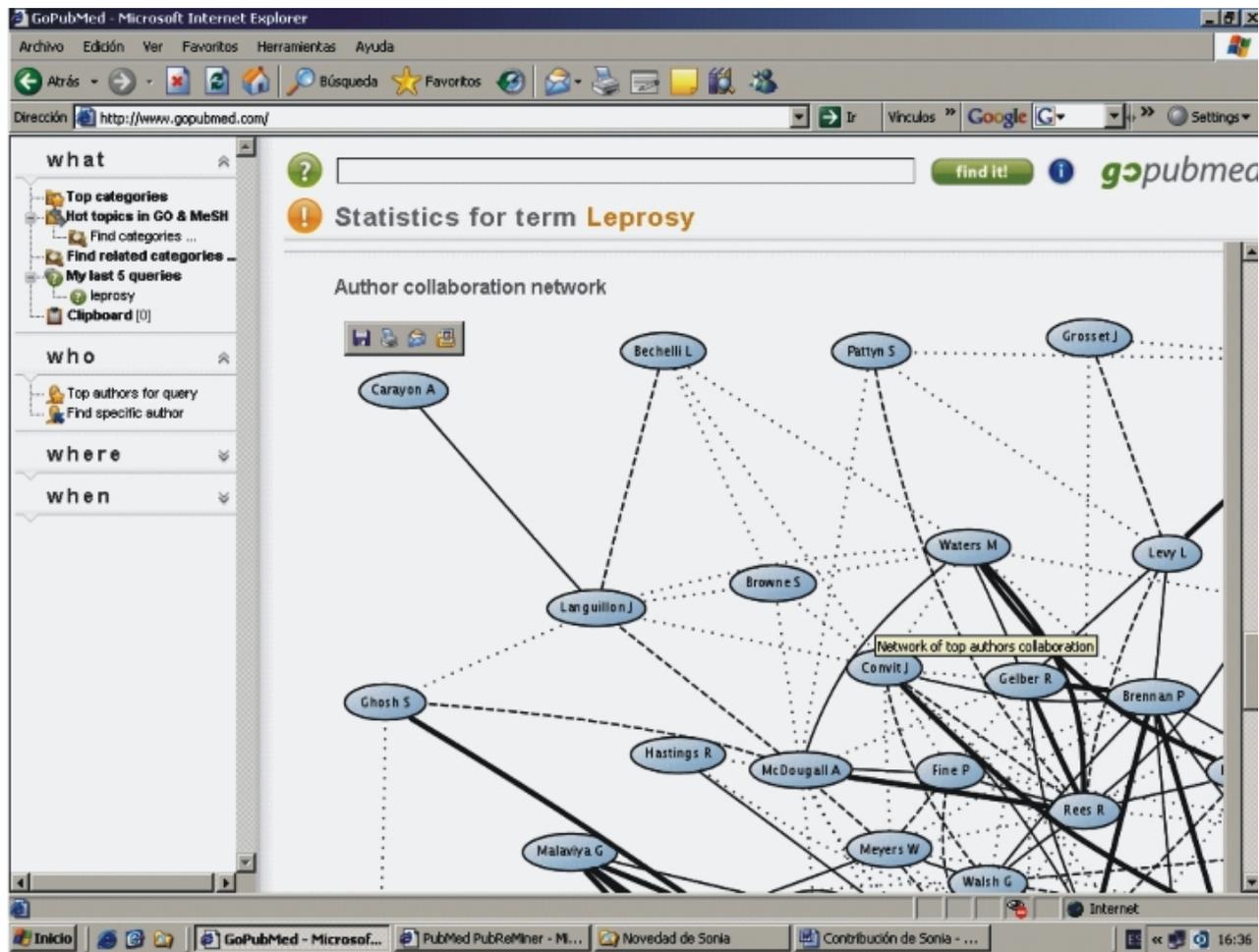


Fig. 10. Red social de autores en la temática más frecuente.

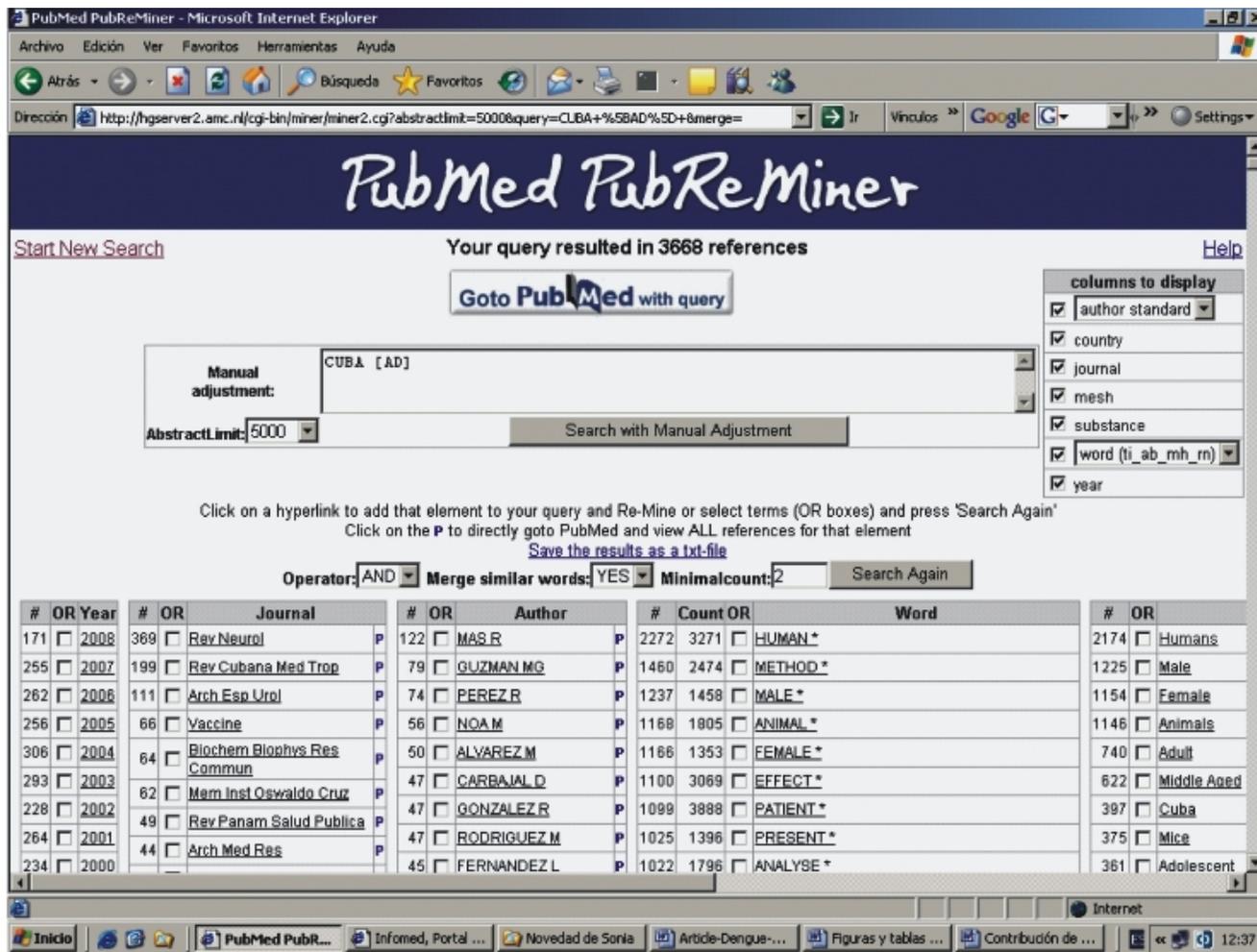


Fig. 15. Análisis de la producción científica de Cuba procesada por PubMed.