

Endo-Mining: herramienta web para la búsqueda automatizada de genes potencialmente relacionados con la endometriosis a través de minería de textos en PubMed

Jorge Vallejo Ortega

22/03/2021

Índice

1	Búsqueda en PubMed	1
2	Exploración de paquetes de minería de textos	3
3	Referencias	3

1 Búsqueda en PubMed

PubMed es un recurso en línea de acceso público y gratuito consistente en una base de datos - en continuo crecimiento - que incluye más de 32 millones de citas y abstracts de literatura biomédica, tanto artículos (*MEDLINE* y *PubMed Central*) como libros (*Bookshelf*). Esta base de datos - en línea desde 1996 - fue desarrollada y sigue siendo mantenida por el Centro Nacional para la Información Biotecnológica (*National Center for Biotechnology Information*, NCBI), que forma parte de la Biblioteca Nacional de Medicina de los E.E.U.U. (*U.S. National Library of Medicine*, NLM) de los Institutos Nacionales de Salud (*National Institutes of Health*, NIH). Esta base de datos está especializada en publicaciones centradas en campos científicos relacionados con la salud y la biomedicina (National Library of Medicine (2021a), National Library of Medicine (2021b)).

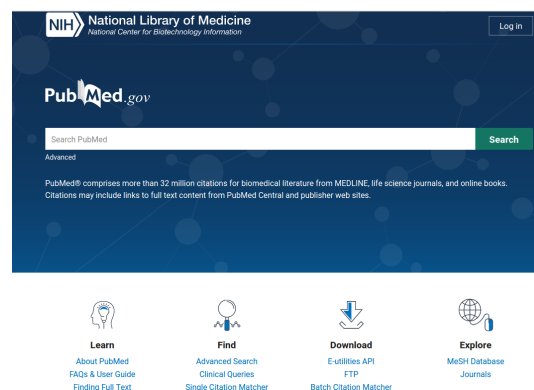


Figure 1: Página principal de PubMed. La barra de búsqueda destaca en medio de la imagen, indicando la finalidad principal de esta página.

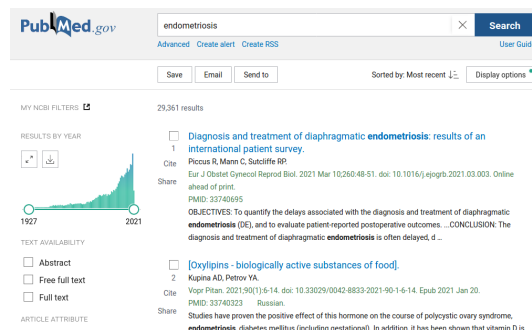


Figure 2: Página de resultados para el término 'endometriosis'. En la parte central se muestran las citas recuperadas por el algoritmo. En el margen izquierdo se nos ofrecen filtros interactivos para refinar la búsqueda.

Para realizar búsquedas es posible utilizar una serie de etiquetas con las que especificar si el término que hemos escrito debe buscarse como parte del título (etiqueta [TI]), el autor ([AU]), la revista ([TA]) o cualquier otro campo dentro de una larga lista que podemos consultar en la sección de ayuda de PubMed¹. También es posible usar los operadores booleanos AND, OR y NOT. Sin embargo, no es necesario emplear las etiquetas ni los operadores, ya que el motor de búsqueda de la página puede crear por sí mismo búsquedas complejas a partir de sólo las palabras clave que introducimos en el campo de búsqueda.

El algoritmo que construye las búsquedas a partir de nuestras palabras clave (*Automatic Term Mapping*) contrasta dichas palabras clave contra diferentes tablas de traducción de términos. En este orden: tabla de traducción de temas, tabla de traducción de revistas, índice de autores e índice de investigadores (colaboradores). Cuando se encuentra una coincidencia para el término o la frase, dicha coincidencia se añade a la búsqueda y no se continúa en la siguiente tabla de traducción.

La tabla de temas relaciona - entre otras cosas - las diferentes formas ortográficas del inglés americano y el británico, formas singulares y plurales, sinónimos, términos fuertemente relacionados, nombres de medicamentos genéricos y sus nombres comerciales, y el vocabulario controlado incluido en el tesauro MeSH (_Medical Subject Headings).

La tabla de revistas contiene y relaciona el título completo de las revistas, sus abreviaciones y sus números ISSN.

El índice de autores y el índice de investigadores contienen el nombre, iniciales y nombre completo de los autores incluidos en la base de datos.

Primero de todo pruebo a realizar una búsqueda en PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). La más básica posible y más general, sin ningún filtro, con la palabra clave "endometriosis". El resultado son 29,361 citas, desde 1927 hasta 2021.

A través del enlace 'Advanced', que se encuentra en la zona superior izquierda, podemos acceder a un constructor de búsquedas que nos facilita hacer búsquedas avanzadas sin necesidad de conocer todas las etiquetas disponibles. En esa misma página podemos consultar nuestro historial de búsquedas recientes y, en éste, cómo el constructor de búsquedas ha traducido nuestra búsqueda simple ('endometriosis') utilizando las tablas de traducción:

Asimismo si usamos los filtros para, por ejemplo, limitar la búsqueda a los artículos publicados durante los últimos diez años, también podemos consultar la estructura de dicha búsqueda:

¹<https://pubmed.ncbi.nlm.nih.gov/help/#search-tags>

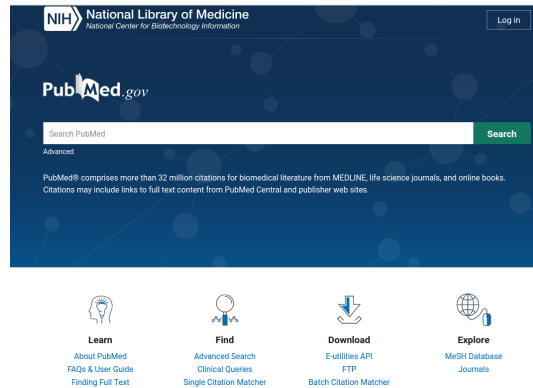


Figure 3: Detalle del historial de búsqueda. De izquierda a derecha muestra la siguiente información: número de la búsqueda en orden cronológico, los términos de búsqueda entrados por el usuario y los términos a los que el algoritmo de mapeo automático los ha traducido, el número de resultados y finalmente la hora a la que se solicitó la búsqueda.

Search	Actions	Details	Query	Results	Time
#2	...	▼	Search: endometriosis Filters: from 2010 - 2021 Sort by: Most Recent ("endometriosis"[MeSH Terms] OR "endometriosis"[All Fields] OR "endometrioses"[All Fields]) AND (2010:2021[pdat]) Translations endometriosis: "endometriosis"[MeSH Terms] OR "endometriosis"[All Fields] OR "endometrioses"[All Fields]	12,865	14:16:50

Figure 4: Detalle del historial de búsqueda. Misma búsqueda del ejemplo anterior, filtrada para obtener como resultado citas de entre los años 2010 y 2021

2 Exploración de paquetes de minería de textos

blablebliblioblu

3 Referencias

- National Library of Medicine. 2021a. “About - PubMed.” Accessed March 9. <https://pubmed.ncbi.nlm.nih.gov/about/>.
- . 2021b. “MEDLINE Overview.” Accessed March 9. https://www.nlm.nih.gov/medline/medline__overview.html.