

HERRAMIENTAS DE MINERÍA DE TEXTO QUE AUTOMATIZAN Y AYUDAN EN EL PROCESO DE REVISIÓN DE LA BIBLIOGRAFÍA

TEXT MINING TOOLS THAT AUTOMATE AND HELP IN THE BIBLIOGRAPHY REVIEWING PROCESS

Autora: Carmen Rodríguez Otero.

Bibliosaúde. Servizo Galego de Saúde (Sergas).

Contacto : maria.carmen.rodriguez.otero@sergas.es

Fecha de recepción: 28/12/2018

Fecha de aceptación: 31/12/2018

Rodríguez-Otero C. Herramientas de minería de texto que automatizan y ayudan en el proceso de revisión de la bibliografía. *Enferm Dermatol.* 2018;12(35): 53-55. DOI: 10.5281/zenodo.2550517

RESUMEN:

El proceso de revisión bibliográfica es una parte primordial de la investigación, este documento presenta tres sencillas herramientas de minería de texto que automatizan y ayudan en las fases de búsqueda y evaluación de este proceso.

Palabras clave: Minería de datos, Revisión de la literatura como asunto, Bases de datos como asunto.

ABSTRACT:

The bibliographic review process is a fundamental part of the research, this document presents three simple text mining tools that automate and help in the search and evaluation phases of this process.

Key words: Data mining, Review Literature as Topic, Databases as Topic.

CONTENIDO:

La revisión bibliográfica es una etapa esencial que debe ser acometida independientemente

de la investigación que se pretenda realizar, además, debe garantizar la obtención de la información más relevante en el campo de estudio⁽¹⁾, dentro de la ingente cantidad de información disponible y la diversidad de fuentes y formatos que la recogen⁽²⁾. Este proceso se debería afrontar dentro de un marco riguroso y sistemático⁽³⁾ independientemente del producto al que den lugar: trabajos académicos, proyectos de investigación o artículos científicos. Las fases de búsqueda y evaluación de la información⁽⁴⁾ deberían afrontarse desde una metodología sistematizada.

En este contexto de recuperación y evaluación de la información, las aplicaciones de minería de texto resultan útiles tanto en las tareas de búsqueda con la identificación de conceptos y selección de términos como en la fase de evaluación en la selección de los estudios.

Las aplicaciones de *minería de texto* están diseñadas para analizar documentos y ayudar a priorizar lo más relevante en función de la pregunta de estudio, de acuerdo a las reglas con que se programen. A continuación se presentan 3 de estas herramientas que ayudan en las fases de búsqueda y selección de los estudios.

Búsqueda de estudios:

En la planificación de la búsqueda es necesario definir e identificar los conceptos y seleccionar los términos que se utilizarán en las bases de datos.

Por ello, utilizar “MeSH on Demand” (Imagen 1) (<https://meshb.nlm.nih.gov/MeSHonDemand>)

puede ser de ayuda en la identificación de términos a partir de un resumen del tema o pregunta de estudio. Esta herramienta, elaborada por la Biblioteca Nacional de Medicina de EE.UU. (NLM), analiza el texto que el usuario facilita y a través del Medical Text Indexer (MTI) devuelve términos relevantes del Medical Subject Heading (MeSH) proporcionando los descriptores para realizar la búsqueda bibliográfica en PubMed, incluso genera una estrategia de búsqueda automática a partir del texto introducido en la herramienta. Breve tutorial de uso en:

<https://mariamontanavivas.wordpress.com/tag/mesh-on-demand/>

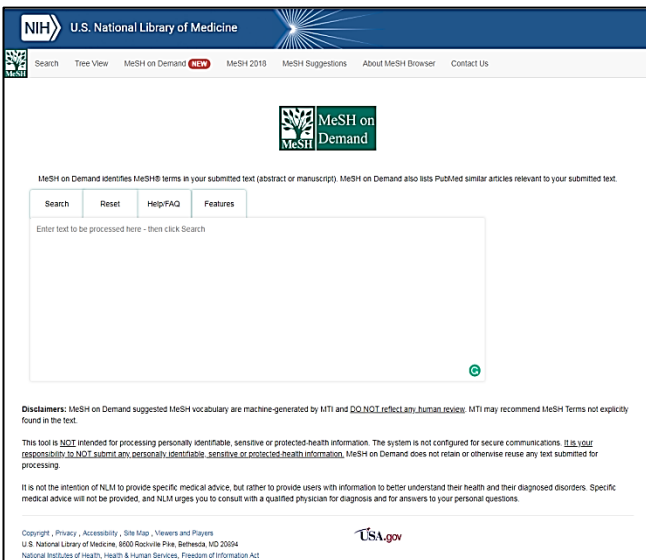


Imagen 1: Página principal de MeSH on Demand.

Ecuación (estrategia) de búsqueda:

Una vez seleccionados los términos deben combinarse mediante operadores booleanos para elaborar la ecuación de búsqueda que se

usará en cada base de datos. Este puede ser un proceso complejo porque exige conocer las características de cada base de datos, en esta tarea son útiles herramientas que traducen las ecuaciones de búsqueda al formato de cada base de datos, como:

“Polyglot Search” (<https://crebp.github.io/sra-polyglot/>) (Imagen 2), elaborada por Bond University Centre for Research in Evidence-Based Practice, permite convertir ecuaciones de búsqueda a formatos de diferentes bases de datos médicas, por ejemplo, se puede realizar una búsqueda en PubMed y convertirla a otra base de datos como Embase. Está disponible para las siguientes bases de datos: PubMed, Ovid Medline/Ovid Embase, Cochrane Library, Embase, Web of Science, CINAHL, PsycInfo, Scopus, Lexical Tree (JSON), Lexical Tree (Human Readable) y MongoDB Query Format.

Para su aplicación basta con pegar la ecuación de búsqueda en el formulario disponible y el resultado aparecerá al desplegar cada una de las bases de datos.

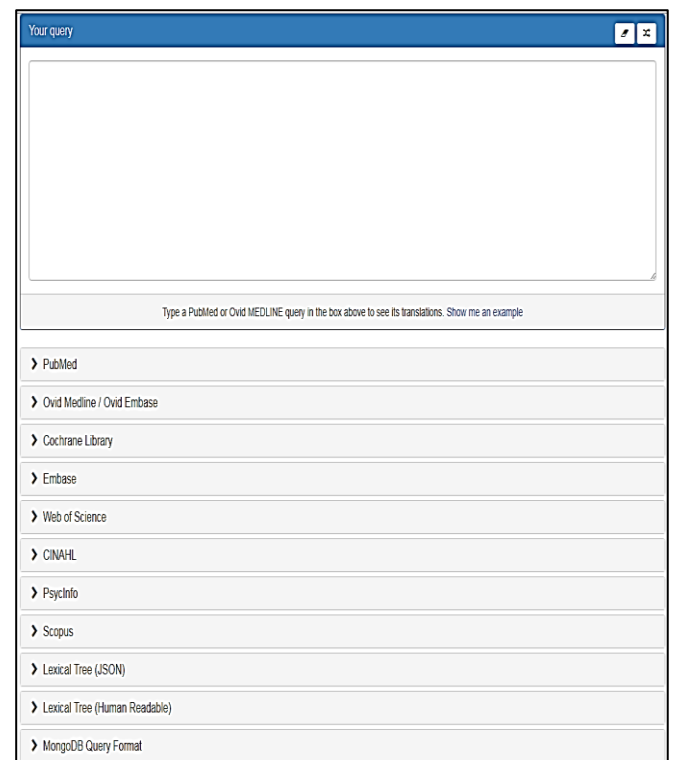


Imagen 2: Página principal de Polyglot.

Más herramientas similares en:

<https://bibliosaude.sergas.gal/DXerais/922/biblioteca%20herramientas%20conversi%C3%B3n%20estrategias%20b%C3%BAqueda.pdf>

Selección de los estudios:

La selección e identificación de los estudios es una de las tareas que más tiempo requieren. Con frecuencia el autor tiene que enfrentarse a largas listas de referencias bibliográficas en un procesador de texto y realizar el cribado manualmente. Diversos estudios evaluaron la aplicación de la minería de textos en esta tarea (5,6), entre las aplicaciones analizadas se encuentra:

“Rayyan QCRI” (<https://rayyan.qcri.org/welcome>)

(Imagen 2) que fue desarrollada por el Qatar Computing Research Institute para agilizar el cribado inicial de títulos y resúmenes utilizando un proceso semiautomático⁽⁷⁾. Esta herramienta gratuita, después de un registro previo, permite subir a la plataforma los resultados de las búsquedas en la mayoría de los formatos de gestión de referencias (Bibtext, Endnote, RefMan, csv...). Una vez subidas las referencias se puede filtrar, etiquetar, exportar, navegar por ellas o subir los textos completos. Además se mostrarán en el escritorio todas las revisiones en las que el usuario participe, tanto como creador o propietario como colaborador, traductor o lector, posibilitando que varios colegas trabajen con el mismo conjunto de referencias para incluir o excluir estudios. A medida que se etiquetan los estudios como incluidos o excluidos, la aplicación “aprende” de estas decisiones y una vez etiquetados suficientes artículos Rayyan empezará a hacer sugerencias con un sistema de valoración propio.

Está disponible como aplicación web y aplicación para dispositivos móviles. Se puede consultar un breve tutorial de uso en:

<https://www.youtube.com/embed/irAOQgzFMs4>

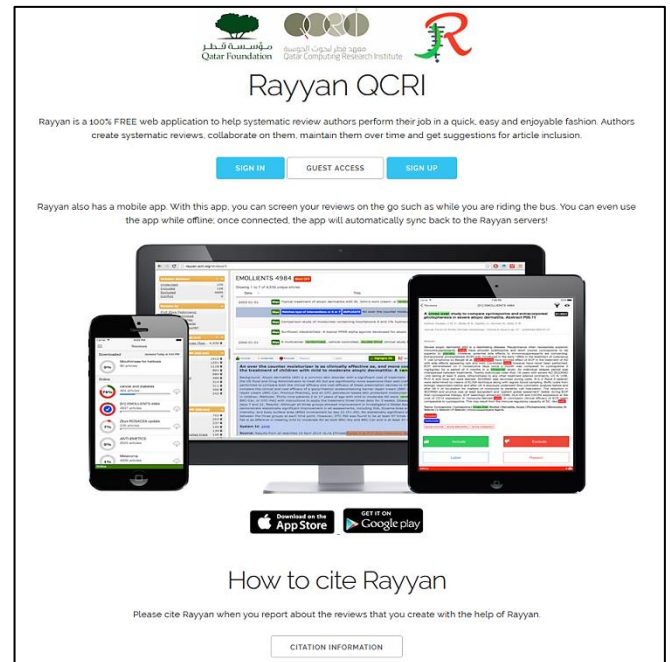


Imagen 3 Página principal de Rayyan QCRI.

BIBLIOGRAFÍA:

1. Gómez-Luna E, Fernando-Navas D, Aponte-Mayor G, Betancourt-Buitrago LA. Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización. Dyna [Internet]. 2014;81(184):158-63.
2. Aleixandre-Benavent R, Alcaide GG, de Dios JG, Alonso-Arroyo A. Fuentes de información bibliográfica (I). Fundamentos para la realización de búsquedas bibliográficas. Acta Pediátrica Esp. 2011;69(3):131-6.
3. Codina L. Revisión bibliográfica sistematizada: procedimientos generales y Framework para ciencias humanas y sociales. Barcelona: Universitat Pompeu Fabra, Departamento de Comunicación, Máster Universitario en Comunicación Social; 2018.
4. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J. 2009;26(2):91-108.
5. Khabsa M, Elmagarmid A, Ilyas I, Hammady H, Ouzzani M. Learning to identify relevant studies for systematic reviews using random forest and external information. Mach Learn. 2015;1-18.
6. Olofsson H, Brolund A, Hellberg C, Silverstein R, Stenström K, Österberg M, et al. Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. Res Synth Methods. 2017;8(3):275-80.
7. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.