

Optimización de algoritmos de filtro en bases de datos bibliográficas para la mejora en la recuperación de información científica.

Galo Enrique Estupiñán Vera

galo.estupinanv@ug.edu.ec

<https://orcid.org/0000-0002-2356-2271>

Afiliación: Universidad de Guayaquil

País: Ecuador

Artículo recibido: 22 agosto 2024

Aceptado para publicación: 19 noviembre 2024

RESUMEN

La eficacia en la búsqueda de información científica es crucial para la investigación académica. Este estudio evalúa el rendimiento de filtros avanzados en las bases de datos Scopus y Web of Science, mediante el uso de ecuaciones de búsqueda específicas. Se analiza cómo la implementación de operadores y field codes afecta la precisión y relevancia de los documentos recuperados. Se destaca el papel de los filtros Open Access y el uso de inglés en los field codes TITLE-ABS-KEY para obtener resultados más precisos.

Palabras clave: *investigación; filtros; ecuación; resultados.*

Optimization of Filter Algorithms in Bibliographic Databases to Improve the Retrieval of Scientific Information

ABSTRACT

Efficiency in the search for scientific information is crucial for academic research. This study evaluates the performance of advanced filters on Scopus and Web of Science databases, using specific search equations. It analyzes how the implementation of operators and field codes affects the precision and relevance of the recovered documents. The role of Open Access filters and the use of English in the



TITLE-ABS-KEY field codes is highlighted to obtain more precise results.

Keywords: *research, filters, equation, results*

INTRODUCCIÓN

La investigación científica exige una recuperación eficiente y precisa de información relevante. Con el avance de las bases de datos digitales, el uso de filtros avanzados y algoritmos personalizados se ha vuelto indispensable. Este estudio surge de la necesidad de evaluar y optimizar estos filtros, particularmente en *Scopus* y *Web of Science*, para maximizar la relevancia y minimizar el tiempo de búsqueda. Siguiendo la metodología propuesta por Guillermina Moncada-Hernández (2014), este trabajo investiga las etapas críticas para una búsqueda efectiva y propone las siguientes mejoras algorítmicas:

1. Acercamiento al tema: Realizar búsquedas referentes al tema en cualquier recurso que permita abarcar más conocimiento general e identificar conceptos y palabras únicas resumidas en 1 o 2 frases.
2. Planteamiento de la pregunta: Formular preguntas jerárquicas que van de lo general a lo particular con el uso de palabras en idioma inglés y tesauros, además de hacer relaciones entre ellas, véase tabla 1.
3. Construcción de la estrategia: Uso de herramientas de búsqueda avanzada en BD para identificar componentes como los campos a utilizar, véase tabla 2 y 3.
4. Elección de la fuente de información: Elección del recurso bibliográficos apropiado para el tema en interés, pues no todos los BD se dirigen para un área como la salud o agrónoma.
5. Refinar la búsqueda: Los resultados pueden llegar a ser muy bajos o altos, se debe ampliar o limitar los puntos de la estrategia aplicada para obtener una cantidad manejable.
6. Organización, administración de la información: Organizar y documentar apropiadamente los resultados obtenidos.

Los investigadores requieren filtros para encontrar información relevante, partiendo desde fuentes confiables disponibles en BD como *Scopus* y *Web of Science* WOS, estos predisponen de un sistema de filtros básicos y avanzados que además requiere criterio de búsqueda del investigador.

Las herramientas de búsqueda devuelven un conjunto de resultados gracias a que realizan la comparación de criterios bibliográficos entre los objetivos de búsqueda con la información de los campos

de metadatos seleccionados existentes en la base de datos (Medrano, 2018), a pesar de que las herramientas mantienen similitudes por las funciones que poseen cada una de los *Repositorios Institucionales* (RI), por ejemplo: **Operators – Operadores**

En general, RI incorporan el uso de operadores booleanos comunes (AND, OR, AND NOT) para la combinación de términos que se usen para una búsqueda más precisa.

Tabla 1

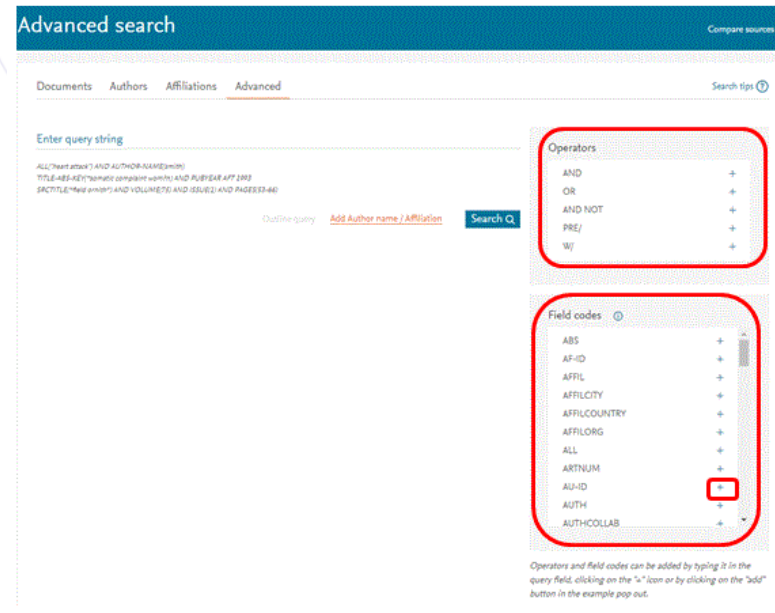
Función operadores booleanos

OPERADOR BOOLEANO	RESULTADOS
OR	Al menos un término aparece
AND	Obligatoriamente todos los términos aparecen
AND NOT	Exclusión de un término

Scopus

Imagen 1

Sección búsqueda avanzada Scopus. Field codes y operators.



Nota: Imagen tomada de internet

Scopus es una plataforma web que contiene una base de datos de resúmenes y citas completas, multidisciplinaria y confiable que permite encontrar investigaciones relevantes y autorizada, acceder a

datos, métrica y herramientas analíticas actualizadas, dispone de 91+ millones de registros, 27950 títulos de series activos y más de 292000 libros (Elsevier, 2023).

En el repositorio de *Scopus* existe una cantidad alta de criterios de búsqueda en su sección “búsqueda avanzada”, en la línea de código query string es posible la visualización del código que se arma a partir de los campos y palabras claves que se integran y se utilizan para luego realizar una búsqueda con las especificaciones descritas delimitando al mínimo el número de respuestas.

Field codes - Códigos de campo

Para alimentar la especificación de búsqueda existen alrededor de 60 field codes disponibles para un término concreto, acorde a *Scopus* (2018) entre los field codes más frecuentes en las búsquedas están:

Tabla 2
Códigos de búsqueda frecuentes

CÓDIGO	DESCRIPCIÓN
TITLE	Título del documento
TITLE-ABS-KEY	Combinación de campos de título del documento, resumen y palabras clave
AF-ID	Afiliación ID de organismos institucionales de Scopus
AU-ID	Identificador de autores de <i>Scopus</i>
ORCID	Número de 16 dígitos para identificar individualmente a autores
KEY	Campo combinado que busca en los campos <i>AUTHKEY</i> , <i>INDEXTERMS</i> , <i>TRADENAME</i> y <i>CHEMNAME</i>
ISSN	Número de identificación único asignado a todas las publicaciones en serie
SUBJAREA	Cuatro categorías (Ciencias de la Salud, Ciencias de la Vida, Ciencias Físicas y Ciencias Sociales)

Nota: Obtenido de página oficial de *Scopus*. Tabla elaborada por autores

Exactitud de frases

Aquellas palabras que se encuentren dentro de comillas, por ejemplo, TITLE-ABS-KEY (“inteligencia emocional”) arrojará resultados en la que inteligencia emocional aparezcan juntas en los resultados conseguidos en el campo combinado de título, resumen y palabras clave de los documentos.

Por otra parte, si no se hace uso de comillas, por ejemplo, TITLE-ABS-KEY (inteligencia emocional), los resultados serán más cuantiosos ya que inteligencia emocional puede aparecer en cualquier orden en el campo especificado (Scopus: Access and use Support Center, 2018).

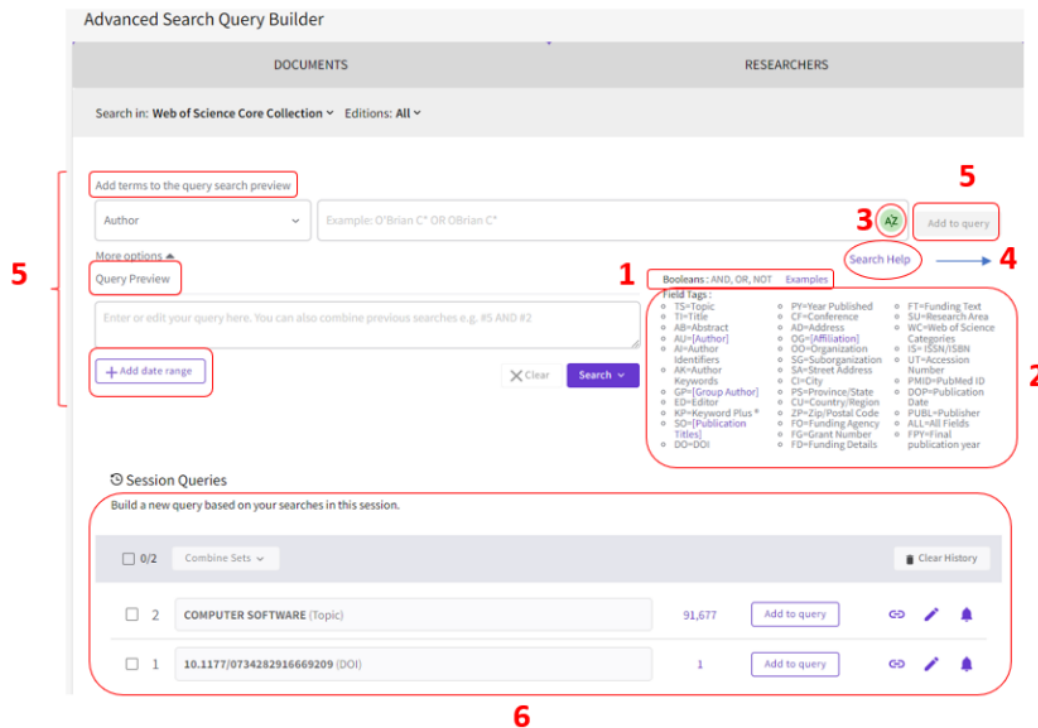


Web of Science WOS

Es una plataforma web que desde 1945 dispone herramientas de análisis que valora la calidad científica. Es suministrada por Clarivate Analytics y abarca 254 disciplinas académicas entre las que se encuentran Ciencias Naturales, Tecnología, Ciencias Sociales, Artes, etc. A través de la recolección de referencias proporciona una data completa a la comunidad investigadora (Chen et al., 2023).

Imagen 2

Composición de búsqueda avanzada en Web of Science: 1. Operadores booleanos, 2. Field labels, 3. Índice términos exactos, 4. Ayuda de búsqueda, 5. Constructora de ecuaciones (query), 6. Resultados basados en el query



Nota: Imagen tomada de internet

Field labels – Etiquetas de campo

La búsqueda avanzada en términos de *Web of Science* incluye alrededor de 40 field labels que permitirán la correcta filtración de información, de acuerdo a Codina (2020) entre las field labels que se usan con más frecuencia, tenemos:

Tabla 3



Códigos de búsqueda

CÓDIGO	DESCRIPCIÓN	CÓDIGO	DESCRIPCIÓN
TS	Tema	SO	Nombre de publicación
TI	Título	DO	DOI
AU	Autor	PY	Año de publicación
AI	Identificadores de autores	AD	Dirección
GP	Autoría conjunta	SU	Área de investigación
ED	Editor	IS	ISSN (Número Internacional Normalizado de Publicaciones Seriadadas) / ISBN (Número Estándar Internacional de Libros)

Nota: Obtenido de guía de utilización de WOS. Tabla elaborada por autores

MATERIALES Y MÉTODOS

El método se remonta en la revisión de patrones en trabajos de titulación en áreas de industria alimenticia, cosmética y farmacéutica de donde se obtienen algoritmos de búsqueda que, mediante un análisis diferencial de la información, permite identificar los indicadores bibliométricos. Además de documentación complementaria que fundamente los términos comprendidos. La revisión literaria de los documentos mencionados da paso a la presente investigación bibliográfica que contiene las diversas fuentes de información mencionadas.

Diseño

La metodología se basa en recopilar información bibliográfica descrita en los trabajos de titulación por medio de cuadros que interpreten un código binario para reflejar la existencia o no del uso de campos que se relacionan directamente con la cantidad de resultados obtenidos.

Fuente de información

La información se basa en trabajos de titulación creados por estudiantes de la facultad de Ingeniería Industrial de la universidad de Guayaquil las cuales consideran la investigación y experimentación de análisis bibliométricos en los principales repositorios virtuales de *Scopus* y *Web of Science* aplicadas a los sectores industriales alimenticios (Mendoza Romero & Luna Sánchez, 2023), farmacéuticos y cosméticos (Martínez Vera & Andrade Merchán, 2023)

METODOLOGÍA

1. Identificar información de la fuente de información

Con los documentos seleccionados como fuente de información se identifica las secciones del capítulo II correspondiente a ecuaciones de búsqueda de BD *Scopus* y *WOS* utilizadas en el desarrollo de las investigaciones, se identifica la totalidad de campos utilizados como una función de filtro, se analiza coincidencias en su estructura para limitar los campos de la presente investigación.

2. Algoritmos en key y arreglos

Para cada ecuación de búsqueda se considera la cantidad de arreglos, estos se conforman por la unión de dos o más *keys* separadas con operadores booleanos AND, la cantidad existente de arreglos está directamente relacionado con la separación del operador booleano OR entre sí, véase en tabla 1, por ejemplo: TITLE-ABS-KEY (inteligencia AND artificial) OR TITLE-ABS-KEY (inteligencia AND artificial AND emocional), existen 2 arreglos con 3 *key* en total, tanto la palabra **inteligencia** como **artificial** se repiten en ambos arreglos, se contabilizan como una sola palabra utilizada en la ecuación de búsqueda, véase resultados en la tabla 6.

2.1. Parametrizar datos

Parametrizar es derivar el análisis del objeto y/o campo de estudio en la investigación con elementos medibles u observables que permitan la valoración o emisión de juicios de valor acerca del estado, nivel o desarrollo del fenómeno o proceso investigado. (Añorga Morales, 2008)

2.2. Segmentar tipos de variables

Los campos se dividen en función al tipo de variable, los datos ausentes en cada registro se completan con: N/A: not available (no disponible) · not applicable (no corresponde en el caso)

Etiquetas de campo (Variable cualitativa)

- Rango de años
- Tipo de archivo

- Subárea
- Lenguaje
- Open Access

Arreglos (Variable cuantitativa)

- Key
- Arreglos and key
- Exclude key
- Arreglos and exclude key

3. Aglomeración de datos

En la tabla 4 se muestra en cada fila una ecuación de búsqueda, cada una específica a qué industria pertenece y en qué BD se ejecutó, además de información correspondiente a las cinco variables cualitativas identificadas y los resultados obtenidos ordenados de manera descendente.

La tabla 6 muestra en cada fila una ecuación de búsqueda, donde se especifica la industria y en qué BD se ejecutó, además de información correspondiente al código manipulado y las variables cuantitativas identificadas relacionadas a los resultados de búsqueda de manera descendente.

4. Transformación binaria

Las variables cualitativas identificadas en la tabla 4 se transformaron a una forma categórica dicotómica o binaria los cuales se presentan en la tabla 5, debido a que por las características de ser variables cualitativas no son aptas para correlacionar los resultados de las búsquedas. Los modelos de regresión logística binaria, como se ha visto, son aquellos en los que la variable dependiente será una variable dicotómica que se codificará como 0 o 1 “ausencia” y “presencia” respectivamente (Rojo Abuín, 2007).

RESULTADOS Y DISCUSIÓN



Tablas

Tabla 4
Aglomeración de datos

#	INDUSTRIA	BUSCADOR	RANGO DE AÑOS	TIPO DE ARCHIVO	SUBAREA	LENGUAJE	OPEN ACCES	RESULTA DOS
1	Alimentos	SCOPUS	2020-2024	AR	ENVI & ENGI	ESPAÑOL & INGLÉS	SI	2223
2	Alimentos	WOS	N/A	N/A	N/A	N/A	N/A	12251
3	Cosmética	SCOPUS	N/A	N/A	N/A	N/A	N/A	8517
4	Cosmética	SCOPUS	N/A	N/A	N/A	N/A	N/A	3972
5	Cosmética	SCOPUS	N/A	N/A	N/A	N/A	N/A	74
6	Cosmética	SCOPUS	2023-2016 & 2013-2012	N/A	N/A	N/A	N/A	63
7	Cosmética	SCOPUS	2023-2016 & 2013-2012	N/A	N/A	N/A	SI	20
8	Cosmética	SCOPUS	2023-2016 & 2013-2012	N/A	N/A	INGLES	SI	20
9	Cosmética	WOS	N/A	N/A	N/A	N/A	N/A	109174
10	Cosmética	WOS	N/A	N/A	N/A	N/A	N/A	109010
11	Cosmética	WOS	N/A	N/A	N/A	N/A	N/A	62861
12	Cosmética	WOS	2023-2012	N/A	N/A	N/A	N/A	52529
13	Cosmética	WOS	2023-2012	N/A	N/A	N/A	SI	26238
14	Cosmética	WOS	2023-2012	N/A	N/A	INGLES	SI	25263
15	Cosmética	WOS	2023-2012	N/A	N/A	INGLES	SI	12
16	Farmacéutica	SCOPUS	N/A	N/A	N/A	N/A	N/A	1184
17	Farmacéutica	SCOPUS	N/A	N/A	N/A	N/A	N/A	483
18	Farmacéutica	SCOPUS	N/A	N/A	N/A	N/A	N/A	483

1	Farmacéutica	SCOPUS	N/A	AR	FHAR & MEDI	INGLES	N/A	267
9								
2	Farmacéutica	SCOPUS	N/A	AR	N/A	N/A	N/A	257
0								
2	Farmacéutica	SCOPUS	N/A	N/A	N/A	N/A	N/A	257
1								
2	Farmacéutica	SCOPUS	N/A	N/A	N/A	N/A	N/A	256
2								
2	Farmacéutica	SCOPUS	N/A	AR	N/A	INGLES	N/A	245
3								
2	Farmacéutica	SCOPUS	N/A	AR	N/A	INGLES	N/A	245
4								
2	Farmacéutica	SCOPUS	N/A	AR	FHAR	N/A	N/A	89
5								
2	Farmacéutica	SCOPUS	N/A	AR	FHAR	INGLES	N/A	25
6								
2	Farmacéutica	SCOPUS	N/A	AR	FHAR	INGLES	N/A	25
7								

Nota: Datos obtenidos de trabajos de titulación. Tabla elaborada por autores.

En la tabla 4 se recopila información que refleja la aglomeración de datos correspondientes a los campos identificados como variables cualitativas de los documentos principales que representa cada industria realizados en buscadores de BD, posteriormente, en la tabla 5 dichas variables se transforman a códigos binarios que ayuda de manera visual en la gráfica 2 a contrastar la existencia en cada una de las ecuaciones de búsqueda experimentadas si se dio uso o no de las variables cualitativas, tomando en cuenta la cantidad de resultados de respuesta.

Tabla 5

Transformación binaria

#	INDUSTRIA	BUSCADOR	RANGO DE AÑOS	TIPO DE ARCHIVO	SUBAREA	LENGUAJE	OPEN ACCES	RESULTADOS
1	Alimentos	SCOPUS	1	1	1	1	1	2223
2	Alimentos	WOS	0	0	0	0	0	12251
3	Cosmética	SCOPUS	0	0	0	0	0	8517
4	Cosmética	SCOPUS	0	0	0	0	0	3972
5	Cosmética	SCOPUS	0	0	0	0	0	74

6	Cosmética	SCOPUS	1	0	0	0	0	63
7	Cosmética	SCOPUS	1	0	0	0	1	20
8	Cosmética	SCOPUS	1	0	0	1	1	20
9	Cosmética	WOS	0	0	0	0	0	109174
10	Cosmética	WOS	0	0	0	0	0	109010
11	Cosmética	WOS	0	0	0	0	0	62861
12	Cosmética	WOS	1	0	0	0	0	52529
13	Cosmética	WOS	1	0	0	0	1	26238
14	Cosmética	WOS	1	0	0	1	1	25263
15	Cosmética	WOS	1	0	0	1	1	12
16	Farmacéutica	SCOPUS	0	0	0	0	0	1184
17	Farmacéutica	SCOPUS	0	0	0	0	0	483
18	Farmacéutica	SCOPUS	0	0	0	0	0	483
19	Farmacéutica	SCOPUS	0	1	1	1	0	267
20	Farmacéutica	SCOPUS	0	0	0	0	0	257
21	Farmacéutica	SCOPUS	0	0	0	0	0	257
22	Farmacéutica	SCOPUS	0	0	0	0	0	256
23	Farmacéutica	SCOPUS	0	0	0	1	0	245
24	Farmacéutica	SCOPUS	0	0	0	1	0	245
25	Farmacéutica	SCOPUS	0	1	1	0	0	89
26	Farmacéutica	SCOPUS	0	1	1	1	0	25
27	Farmacéutica	SCOPUS	0	1	1	1	0	25

Nota: Datos basados en tabla 4. Tabla elaborada por autores.

Otro punto importante es la identificación y contabilización de arreglos de palabras claves y palabras claves excluidas, en la tabla 6 los resultados muestran un énfasis en el uso de campos como ALL y TITLE-ABS-KEY para realizar búsquedas y el impacto que estos tienen.

Tabla 6
Algoritmos key y arreglos

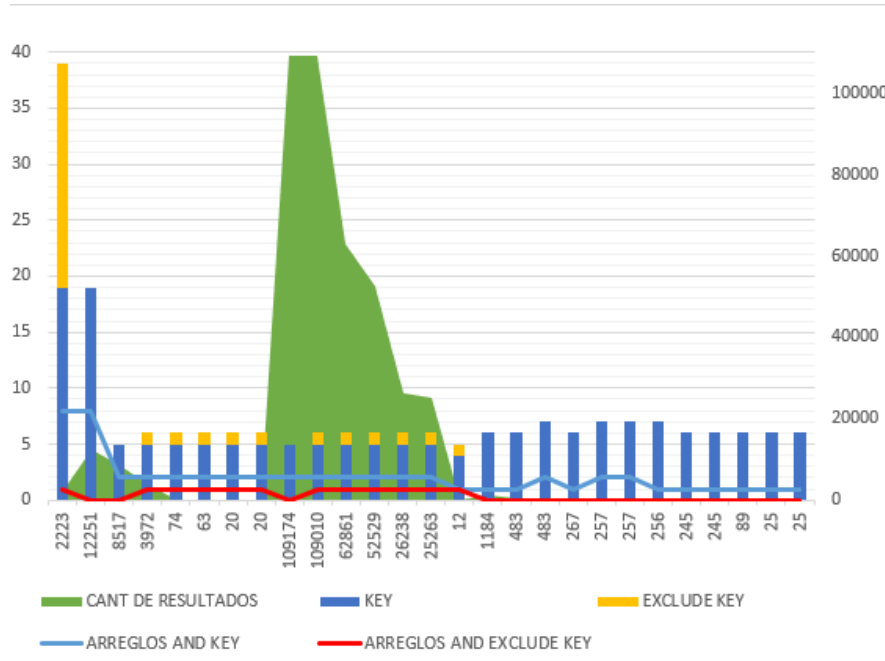
#	INDUSTRIA	BUSCADOR	CÓDIGO	KEY	#ARREGLOS AND KEY	EXCLUDE KEY	#ARREGLOS AND EXCLUDE KEY	RESULTADOS
1	Alimentos	SCOPUS	ALL	19	8	20	1	2223
2	Alimentos	WOS	ALL	19	8	0	0	12251
3	Cosmética	SCOPUS	ALL	5	2	0	0	8517
4	Cosmética	SCOPUS	ALL	5	2	1	1	3972
5	Cosmética	SCOPUS	TITLE-ABS-KEY	5	2	1	1	74
6	Cosmética	SCOPUS	TITLE-ABS-KEY	5	2	1	1	63
7	Cosmética	SCOPUS	TITLE-ABS-KEY	5	2	1	1	20
8	Cosmética	SCOPUS	TITLE-ABS-KEY	5	2	1	1	20
9	Cosmética	WOS	ALL	5	2	0	0	109174
10	Cosmética	WOS	ALL	5	2	1	1	109010
11	Cosmética	WOS	TITLE-ABS-KEY	5	2	1	1	62861
12	Cosmética	WOS	TITLE-ABS-KEY	5	2	1	1	52529
13	Cosmética	WOS	TITLE-ABS-KEY	5	2	1	1	26238
14	Cosmética	WOS	TITLE-ABS-KEY	5	2	1	1	25263
15	Cosmética	WOS	TITLE-ABS-KEY	4	1	1	1	12
16	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	1184
17	Farmacéutica	SCOPUS	ALL	6	1	0	0	483
18	Farmacéutica	SCOPUS	TITLE-ABS-KEY	7	2	0	0	483
19	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	267
20	Farmacéutica	SCOPUS	TITLE-ABS-KEY	7	2	0	0	257
21	Farmacéutica	SCOPUS	TITLE-ABS-KEY	7	2	0	0	257
22	Farmacéutica	SCOPUS	TITLE-ABS-KEY	7	1	0	0	256
23	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	245
24	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	245
25	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	89
26	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	25
27	Farmacéutica	SCOPUS	TITLE-ABS-KEY	6	1	0	0	25

Nota: Datos obtenidos de trabajos de titulación. Tabla elaborada por autores

Gráficas datos generales de campos y resultados obtenidos

Gráfica 1

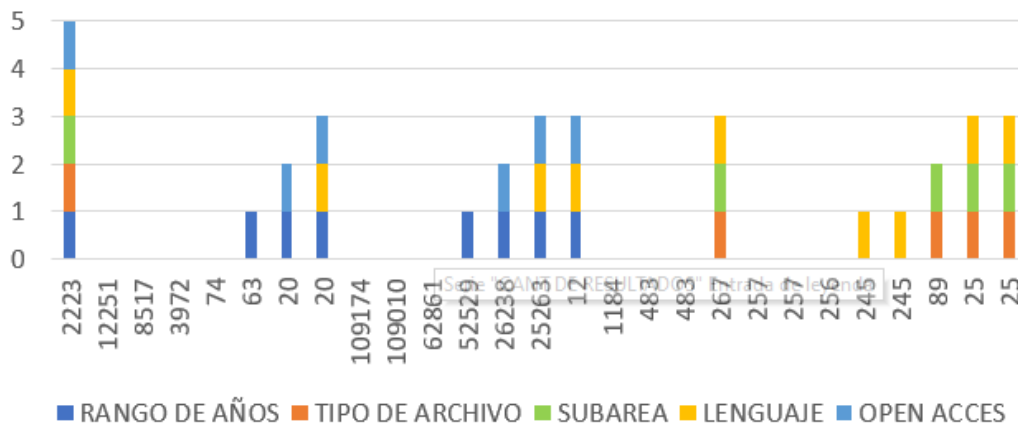
Incidencia de campos usados en búsquedas realizadas en Scopus y Web of Science en cantidad de resultados obtenidos



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

Gráfica 2

Campos usados en búsquedas realizadas en Scopus y Web of Science con relación a la cantidad de resultados



Nota: Datos basados en tabla 5. Gráfica realizada por autores.

En el gráfico 1 se identifica que las búsquedas que han tenido menos de 100 resultados están compuestas entre cuatro y cinco palabras claves (KEY), donde al menos existe una palabra clave excluida (EXCLUDE KEY), y su estructura se presenta con máximo dos arreglos de palabras claves y un arreglo para las palabras claves excluidas. Por consiguiente, es posible deducir el alto grado de influencia que conlleva la selección de palabras claves y se sugiere que estas no excedan la cantidad mencionada previamente.

En contraste los casos donde se utilizaron alrededor de 19 palabras claves se determinaron que no se redujo significativamente los resultados, debido a que estuvo conformado por 8 arreglos los cuales amplia la búsqueda y genera ambigüedad en los términos utilizados, denotando una vez más la importancia de la selección de palabras específicas a introducir en el algoritmo.

En el gráfico 2 se refleja el uso de etiquetas de campos que están inmersos en la estructura de los algoritmos, las búsquedas que han tenido menos de cien resultados presentan como máximo tres tipos de filtros en las etiquetas de campo, donde la que tiene mayor recurrencia en los mejores resultados es el rango de año seguido de open Access, lenguaje, subárea y tipo de archivo, siendo este el orden recomendado para aplicar estos filtros según sea requerido, el rango de año queda a criterio del investigador basado en sus objetivos.

La agrupación de los datos inmersos en ambos gráficos produce estar en capacidad de obtener resultados efectivos debido a la estrecha correlación que tienen como filtros de búsqueda.

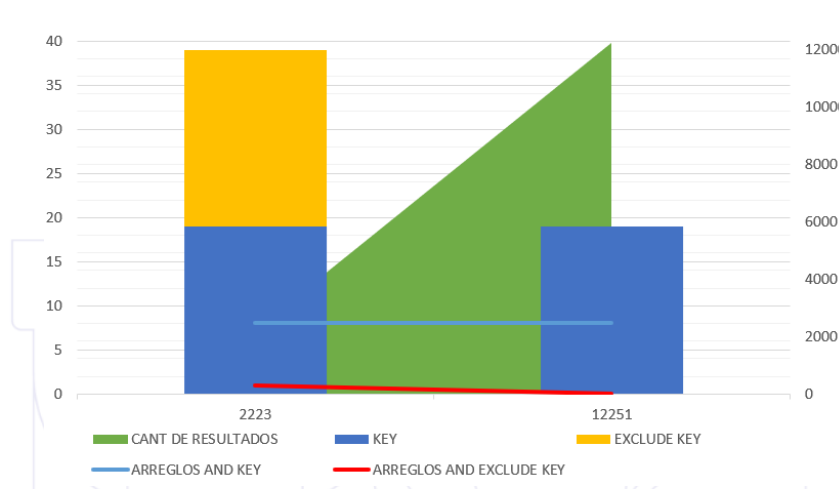
Análisis de la evolución de resultados por tipo de industria

Al momento de realizar alguna investigación siempre existirá diferencias entre las diversas áreas que sea del interés del investigador, no es posible el simple análisis general sin tomar en consideración el comportamiento natural que se obtendrá de cada área, ya sea por su recurrencia en cierto rango de años por las palabras tesauros, entre otros. A continuación, se presenta un análisis por tipo de industria y su comportamiento por cada BD trabajada.

Industria de alimentos

Gráfica 3

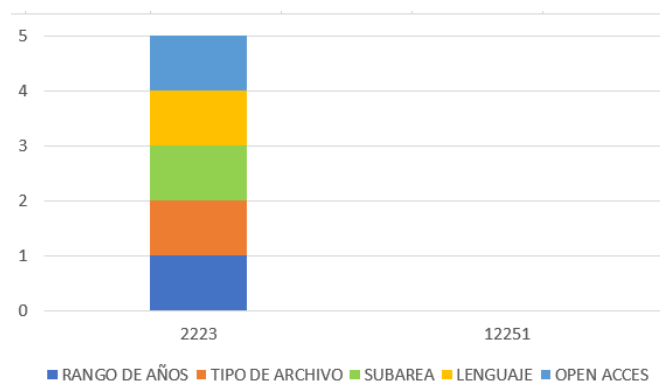
Incidencia de campos usados en búsquedas realizadas en *Scopus* y *Web of Science* con cantidad de resultados obtenidos en industria alimentos



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

Gráfica 4

Campos usados en búsquedas realizadas en *Scopus* y *Web of Science* con relación a la cantidad de resultados en industria alimentos



Nota: Datos basados en tabla 5. Gráfica realizada por autores.

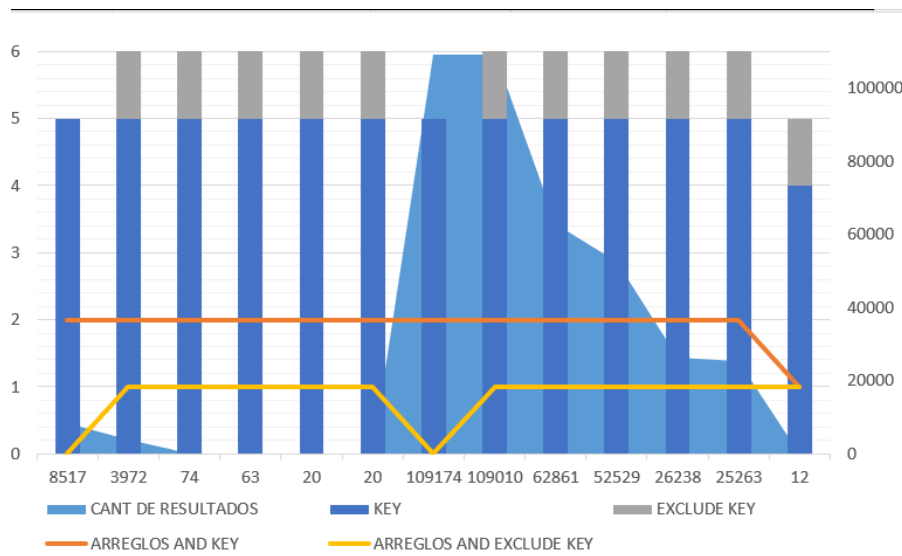
En la gráfica 4, la primera búsqueda con 2223 resultados correspondiente a *SCOPUS* se aprecia la aplicación de cinco etiquetas de campo, suficientes para reducir considerablemente el número de resultados siendo más específicos al propósito de la búsqueda, Sin embargo, en *WOS* se destaca la

ausencia de etiquetas de campo consiguiendo 12251 resultados con temas generales. Aun así, no es posible conseguir una cantidad moderada y manejable.

Industria de cosmética

Gráfica 5

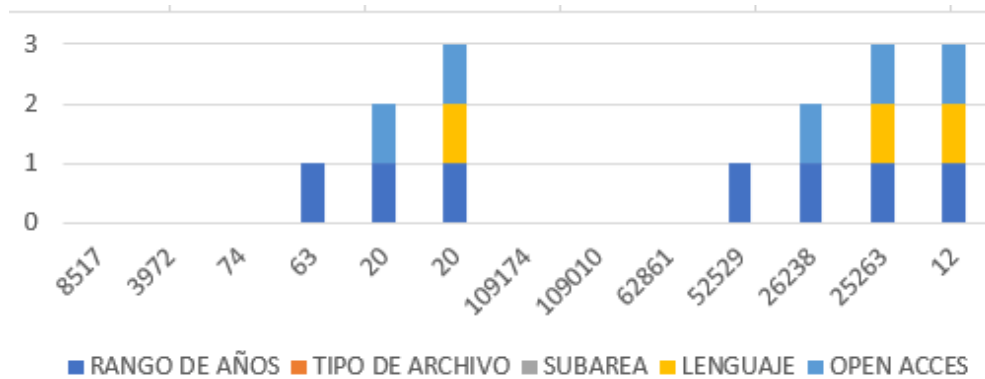
Incidencia de campos usados en búsquedas realizadas en *Scopus* y *Web of Science* con cantidad de resultados obtenidos en industria cosmética



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

Gráfica 6

Campos usados en búsquedas realizadas en *Scopus* y *Web of Science* con relación a la cantidad de resultados en industria cosmética



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

Tomado en cuenta los resultados obtenidos de cada BD, existen comportamientos decisivos para conseguir diferencias enormes de resultados. En la gráfica 5 aplicado en WOS, la ecuación de búsqueda con 12 resultados conseguidos se puede observar el uso de un arreglo de palabras clave, tres etiquetas de campo por lo que el algoritmo realizará búsquedas de documentos que contengan únicamente la aparición de las cinco palabras claves utilizadas y que de los encontrados descarten a aquellos que contengan la única palabra clave excluida, en comparación de la ecuación de búsqueda con 25263 resultados se puede observar el uso de dos arreglos de palabras clave, es decir, que existe más probabilidades de encontrar documentos que contengan ambas especificaciones de arreglos incluidos y que ignoren la única palabra clave excluida.

Ilustración 1

Fórmula para determinar impacto del uso de código TITLE-ABS-KEY

$$\frac{A - B}{A} * 100\%$$

A = resultado mayor aplicando campo ALL

B = resultado menor aplicando campo TITLE – ABS – KEY

Nota: Elaborado por autores

$$\frac{3972-74}{3972} * 100\% = 98\%, \text{ caso 1 en buscador SCOPUS}$$

$$\frac{109010-62861}{109010} * 100\% = 42,3\%, \text{ caso 2 en buscador WOS}$$

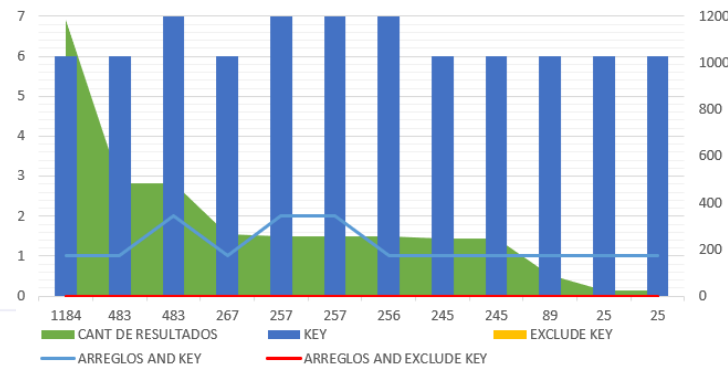
El uso del campo TITLE-ABS-KEY tiene fuerte impacto en los resultados puesto que con su implementación se logra reducir hasta un 98%.

Por otro lado, la reducción de los valores en la BD SCOPUS de 8517 a 20 resultados según la gráfica 6, radica en la cantidad de campos usados teniendo como énfasis el campo de rango de años. En comparación a WOS se observa una reducción de 109174 a 12 resultados debido a la cantidad de campos usados teniendo más impacto el campo de open Access, además de la reducción de dos a un número de arreglos de palabras claves incluidas.

Industria farmacéutica

Gráfica 7

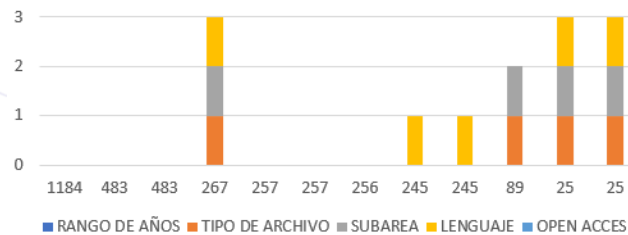
Incidencia de campos usados en búsquedas realizadas en Scopus con cantidad de resultados obtenidos en industria farmacéutica



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

Gráfica 8

Campos usados en búsquedas realizadas en Scopus con relación a la cantidad de resultados en industria farmacéutica



Nota: Datos basados en tabla 6. Gráfica realizada por autores.

En base a gráficas 7 y 8 se puede observar un comportamiento en los resultados de 267 y 25 donde la BD, variables cualitativas y cuantitativas se dieron exactamente el mismo uso, a diferencia de la especificación de la variable cualitativa “subárea” que limita la subárea a una sola opción dando como resultado una diferencia de resultados significativa entre ambas ecuaciones de búsqueda.

$$\frac{483-257}{483} * 100\% = 46.8\%, \text{ caso 1 en buscador SCOPUS}$$

El uso del campo TITLE-ABS-KEY tiene fuerte impacto en los resultados obtenidos, puesto que con su implementación se logra reducir hasta un 46.8%.



Por otro lado, la reducción de los valores en la BD *SCOPUS* de 1184 a 25 resultados, según la gráfica 8, el resultado más alto, al no especificar en ningún campo se da cabida a cantidades de respuesta mucho más amplias y generales. La reducción radica en el uso de tres etiquetas de campo donde tiene mayor relevancia la aplicación del campo “Subárea”.

CONCLUSIONES

Este estudio revela que la efectividad de los filtros avanzados en la recuperación de información no es directamente proporcional al número de campos utilizados o al grado de especificidad de los criterios de búsqueda. En efecto, el uso excesivo de filtros puede conducir a un escenario donde los resultados obtenidos sean mínimos o nulos, lo que subraya la importancia de seleccionar y manipular estratégicamente los datos bibliométricos para alcanzar los objetivos específicos de la investigación.

Los hallazgos sugieren que la utilización de cuatro a cinco términos clave, incluyendo sinónimos relevantes, tiende a optimizar la recuperación de documentos que son pertinentes para los objetivos del investigador, incluso cuando los autores originales pueden no haber empleado las mismas palabras clave designadas en la investigación. Esto destaca la necesidad de una planificación cuidadosa en la selección de palabras clave para abarcar variaciones terminológicas.

La estructuración de los arreglos en las ecuaciones de búsqueda demuestra que, en la mayoría de los casos, la utilización de hasta dos arreglos es suficiente para obtener un conjunto de resultados relevante y manejable, mientras que un solo arreglo puede ser adecuado para búsquedas que requieran una alta especificidad.

Además, se observó que el uso de field codes como TITLE-ABS-KEY es particularmente efectivo, limitando la búsqueda a tres campos esenciales y reduciendo la cantidad de resultados irrelevantes en hasta un 42%. Este resultado es indicativo de la importancia de seleccionar field codes adecuados que correspondan a las necesidades informativas del investigador.

Futuras investigaciones podrían explorar la optimización de algoritmos de búsqueda para balancear la exhaustividad y la precisión, investigar el impacto de diferentes combinaciones de field codes, y desarrollar guías de buenas prácticas para el uso de bases de datos bibliográficas en diversos campos académicos. También sería beneficioso analizar cómo las diferencias en la configuración de búsqueda afectan los resultados en diversas disciplinas y tipos de investigación.

En conclusión, este estudio subraya la necesidad de un enfoque meticuloso y reflexivo en la configuración de búsquedas bibliográficas para maximizar la eficiencia y efectividad en la recuperación de información relevante y precisa.

LISTA DE REFERENCIAS

- Añorga Morales, J. V. I. N. C. S. J. (2008). *La parametrización en la investigación educativa*. 25–32.
- Chen, Y., Koch, T., Zakiyeva, N., Liu, K., Xu, Z., Chen, C. H., Nakano, J., & Honda, K. (2023). Article's scientific prestige: Measuring the impact of individual articles in the web of science. *Journal of Informetrics*, 17(1), 101379. <https://doi.org/10.1016/J.JOI.2023.101379>
- Codina, L. (2020, August 5). *Web of Science: guía de utilización con el procedimiento de las 6 fases*. <https://www.lluiscodina.com/6-fases-web-of-science/>
- Elsevier. (2023). *About Scopus | Abstract and citation database*. <https://beta.elsevier.com/products/scopus?trial=true>
- Guillermina Moncada-Hernández, S. (2014). Cómo realizar una búsqueda de información eficiente. Foco en estudiantes, profesores e investigadores en el área educativa. *Inv Ed Med*, 3(10), 106–115. <http://riem.facmed.unam.mx>
- Martínez Vera, M. R., & Andrade Merchán, M. G. (2023). *Aplicaciones de buscadores informáticos en la investigación de nuevas tecnologías de manejo de desechos en la industria cosmética*. <http://repositorio.ug.edu.ec/handle/redug/67232>
- Medrano, J. F. (2018). Filtrado basado en contenido para artículos académicos en repositorios institucionales. *XXIV Congreso Argentino de Ciencias de La Computación (La Plata, 2018)*, 22–31.



Mendoza Romero, I. S., & Luna Sánchez, D. E. (2023). *Gestión de residuos peligrosos en la industria de alimentos y la socialización de buscadores de internet*.
<http://repositorio.ug.edu.ec/handle/redug/67245>

Rojo Abuín, J. M. (2007). *Regresión con variable dependiente cualitativa*.

Scopus: Access and use Support Center. (2018, January 18). *How can I best use the Advanced search?*
https://service.elsevier.com/app/answers/detail/a_id/11365/c/10546/supporthub/scopus/

