Algoritmos para documentos científicos: pasado y presente

Algorithms for Scientific Documents: Past and Present

Jesús M. Álvarez-Llorente; Vicente P. Guerrero-Bote; Félix De-Moya-Anegón

Cómo citar este artículo:

Álvarez-Llorente, Jesús M.; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2025). "Algorithms for Scientific Documents: Past and Present [Algoritmos para documentos científicos: pasado y presente]". *Infonomy*, 3(4) e25026. *https://doi.org/10.3145/infonomy.25.026*

Artículo recibido: 01-08-2025
Artículo aprobado: 10-09-2025

Jesús M. Álvarez-Llorente
https://orcid.org/0000-0002-4901-3457
https://directorioexit.info/ficha6849
Universidad de Extremadura
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos
Plazuela Ibn Marwam
06071 Badajoz, España
llorente@unex.es

Vicente P. Guerrero-Bote
https://orcid.org/0000-0003-4821-9768
https://directorioexit.info/ficha596
Universidad de Extremadura
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos
Plazuela Ibn Marwam
06071 Badajoz, España
guerrero@unex.es

Félix De-Moya-Anegón
https://orcid.org/0000-0002-0255-8628
https://directorioexit.info/ficha92
SCImago Research Group
18220 Granada, España
felix.moya@scimago.es

Resumen

Este trabajo se presenta como una recopilación de algoritmos de clasificación de la investigación a nivel de artículo como alternativa a las clasificaciones por revistas que se emplean en las grandes bases de datos de ciencia como *Web of Science* o *Scopus*, las cuales causan gran imprecisión en las búsquedas y en la evaluación de la ciencia, ya que utilizando éstas, los artículos no resultan categorizados con fidelidad respecto a su verdadero contenido. En primer lugar hacemos una revisión histórica de las principales ideas planteadas a lo largo de los años desde la misma aparición de las bases de datos, detectando sus contribuciones y sus limitaciones. Los algoritmos de agrupamiento automático y de detección de comunidades han supuesto grandes avances en organización de la ciencia, pero no resultan aplicables como alternativa a la clasificación por revistas. Otros algoritmos no son escalables al conjunto de la ciencia debido a su complejidad, como los basados en redes neuronales o minería de textos. Las propuestas más recientes y prometedoras responden a algoritmos sencillos que, partiendo de la categorización por revistas, reclasifican los artículos en las mismas jerarquías temáticas de las bases de datos, mediante el análisis de simples citas y referencias.

Palabras clave

Algoritmos de clasificación; Clasificaciones a nivel de documento; Clasificaciones; Clasificación de la ciencia; Bases de datos de ciencia; Cienciometría; Citación; Esquemas de clasificación; *ASJC*; Scopus; *Web of Science*.

Abstract

This study offers a comprehensive overview of document-level classification algorithms in scientific research, proposed as an alternative to the journal-based categorizations employed by major bibliographic databases such as Web of Science and Scopus. These journal-driven schemes often introduce significant inaccuracies in both information retrieval and research evaluation, as they fail to categorize articles in accordance with their actual content.

First, we provide a historical review of the main approaches developed since the emergence of scientific databases, highlighting their contributions as well as their limitations. Automatic clustering techniques and community detection algorithms have represented important advances in the organization of scientific knowledge, yet they cannot serve as a practical substitute for journal-based classifications. Other approaches, such as those relying on neural networks or text mining, face scalability issues that prevent their application at the global level of science.

The most recent and promising strategies are built upon simple algorithms that, starting from existing journal categorizations, reclassify articles into the same thematic hierarchies used by bibliographic databases, relying primarily on the analysis of straightforward citation and reference patterns.

Keywords
Classification algorithms; Document-level classifications; Classifications; Science classification; Scientific databases; Scientometrics; Citation; Classification schemes; *ASJC*; *Scopus*; *Web of Science*.

1. Introducción

"La ciencia no es ciencia si no se comparte" es una frase muy repetida en la actualidad, que se atribuye a Nathan Robinson, investigador del *Institut de Ciències del Mar* del *CSIC*, y, aunque con ella Nathan reivindica la importancia de la divulgación de la ciencia por los canales modernos de comunicación, refiriéndose concretamente a las redes sociales, es evidente que el desarrollo de la ciencia sería imposible sin el intercambio de conocimiento entre investigadores mediante los mecanismos tradicionales de publicación científica, principalmente en forma de libros, actas de congresos, etc., y especialmente en revistas científicas que a su vez contienen artículos. Pero, la divulgación no termina en el acto de publicar: otros investigadores necesitan encontrar esa información publicada. En un mundo globalizado con una producción científica anual colosal, las grandes bases de datos científicas, como *Scopus* o *Web of Science* (*WoS*), que indexan esa producción publicada (principalmente las revistas con sus artículos) y la ponen a tiro de búsqueda de los investigadores de todo el mundo, son una herramienta imprescindible para el avance de la ciencia.

En ellas se hace necesario organizar las publicaciones por categorías científicas. La forma más tradicional de hacerlo es categorizar las revistas según su temática. Tanto en la recuperación como en los estudios cienciométricos, es común que esta asignación de categorías a revistas se extienda tal cual a los artículos publicados en ellas, sin tener en cuenta si su contenido real se ajusta en mayor o menor medida a la temática de la revista. Esta forma de organización temática causa problemas en la recuperación de la información debido a una alta imprecisión en las búsquedas. Pero, sobre todos causa problemas en la Cienciometría a la hora de cuantificar la investigación por disciplinas, y también al calcular el impacto de la ciencia, una cuestión vital en la financiación de las investigaciones y los investigadores, ya que son bien sabidas la diferencias en los hábitos de citación en las diferentes disciplinas.

Ya desde los mismos inicios de estas bases de datos se vislumbraba este problema, y por ello se han llevado a cabo numerosos intentos de clasificar los artículos como publicaciones individuales según sus características propias, y no según la revista en la que fueron publicados, a las que a menudo nos referimos también como clasificaciones a nivel de documento (*paper-level*) o clasificaciones *item-by-item*.

Dependiendo del objetivo específico con el que se han diseñado estas clasificaciones podemos encontrar dos grandes formas de proceder:

La primera es la aplicación de técnicas de agrupamiento automático o detección de comunidades, es decir, agrupación de publicaciones con características similares en clústeres o comunidades, desconocidas a priori, que constituyen las categorías temáticas resultantes.

La segunda idea consiste en ir asignando cada publicación a alguna de las categorías temáticas de un esquema de clasificación prestablecido, por ejemplo, el mismo que emplea la base de datos para clasificar las revistas, en lo que podemos denominar un proceso de reclasificación, o, más descriptivamente, (re)clasificación.

La categoría a la que se asigna cada artículo suele decidirse mediante el estudio de las redes de citación (por citación, co-citación o acoplamiento bibliográfico), por minería de textos (principalmente mediante frecuencia de términos), o también, mediante la hibridación de ambas ideas para una mayor precisión.

Los intentos tradicionales de aplicación de estas técnicas han presentado múltiples limitaciones. Por ejemplo, los métodos de agrupamiento acusan problemas de estabilidad, produciendo resultados diferentes cada vez que se inicia el proceso. Además, producen esquemas de clasificación muy diferentes de los empleados por las bases de datos, lo cual introduce un cierto nivel de incoherencia. También son computacionalmente muy costosos, por lo que solo se han podido ensayar sobre pequeños subconjuntos de la producción científica, y algunos no soportan la posibilidad de pertenencia a más de una categoría a la vez, lo cual no refleja la realidad.

Estas dos limitaciones también las podemos padecer en las técnicas de (re)clasificación, especialmente cuando integran minería de textos o las formas más complejas de análisis de las redes de citación (co-citación y acoplamiento bibliográfico).

En definitiva, parece necesario contar con un sistema de clasificación de publicaciones individuales que mejore la precisión de las clasificaciones por revistas, pero superando las limitaciones conocidas, de manera que sea estable, no introduzca grandes desviaciones con respecto a los esquemas de clasificaciones actualmente aceptados por la comunidad científica, que sea aplicable al conjunto de la ciencia, que admita la pertenencia de los trabajos a más de una categoría, y, por supuesto, que tenga en cuenta las diferencias entre disciplinas, una de las premisas para que una clasificación sea realmente útil en el mundo científico actual.

Cuando hablamos de todas estas ideas o técnicas de clasificación, dado el volumen de información que pretendemos procesar, nos estamos refiriendo, por supuesto, a procedimientos automáticos, por lo que puede resultar más adecuado utilizar el término "algoritmo".

Con el auge actual de la inteligencia artificial, la palabra algoritmo ha adquirido una connotación desmerecidamente negativa. Hablamos de “el algoritmo” como un ente incontrolado y misterioso, que nos vigila, dirige nuestra vida y que amenaza incluso con acabar con ella. Pero un algoritmo no es más que una sucesión de instrucciones para completar una tarea, algo tan inocente como puede ser una receta de cocina. Los

ordenadores, encargados de procesar automáticamente toda la información que manejamos, funcionan mediante algoritmos. Un algoritmo es una descripción formal de cómo hay que hacer algo para obtener el resultado que deseamos.

En el presente trabajo haremos una retrospectiva de los principales algoritmos de clasificación a nivel de documento que se han propuesto a lo largo de los últimos años, y cómo han permitido llegar hasta las propuestas más recientes, que analizaremos con mayor detalle.

2. Contexto

WoS y *Scopus* son ciertamente las bases de datos de ciencia más ampliamente aceptadas por la comunidad científica. Ambas clasifican las revistas científicas en categorías. Para la recuperación, así como en muchos estudios bibliométricos, los trabajos heredan las categorías a las que está adscrita la revista en la que se publicaron. Existen diferencias en los esquemas de categorías de ambas bases de datos.

Scopus utiliza el esquema de clasificación llamado *ASJC* (*All Science Journal Classification*) (**Gómez-Crisóstomo**, 2011; **Wang**; **Waltman**, 2016). Consta de 27 áreas temáticas (*subject areas*), una de las cuales es multidisciplinar (*1000 Multidisciplinary*). A ella se asignan las revistas claramente multidisciplinares como *Nature* o *Science*. Las 26 áreas temáticas restantes se subdividen en un total de 311 áreas temáticas específicas o categorías, con la peculiaridad de que cada una de esas 26 áreas contiene una categoría miscelánea (*miscellaneous*) donde incluir las revistas abiertas a temas muy diversos, pero dentro de un área temática o que no se pueden asignar claramente a ninguna de las áreas temáticas específicas.

Por su parte, *WoS* establece en su nivel principal solo 5 áreas de investigación, categorías generales o áreas amplias (Artes y Humanidades, Ciencias de la Vida y Biomedicina, Ciencias Físicas, Ciencias Sociales, y Tecnología). En un segundo nivel se distinguen —actualmente— un total de 254 categorías temáticas o disciplinas, cada una de ellas incluida en alguna de las 5 áreas amplias. Como vemos, no existe un área multidisciplinar, pero sí encontramos una categoría multidisciplinar pura, además, de varias con el apellido multidisciplinar o interdisciplinar (por ejemplo, "*Computer Science*, *Interdisciplinary Applications*" o "*Humanities, Multidisciplinary*").

Ambas clasificaciones se consideran firmemente aceptadas y han sido utilizadas ampliamente para el estudio de la estructura de la ciencia (**Leydesdorff** *et al*., 2010; 2015; **Hassan-Montero** *et al*., 2014), pero, por supuesto, no son las únicas. En **Gläser** *et al.* (2017) se hace una recopilación bastante completa de diferentes sistemas de clasificación propuestos a lo largo de la historia. Y resulta interesante destacar que entre las conclusiones de este trabajo se resalta que uno de los problemas más frecuentes que deben afrontar todos los sistemas de clasificación es el establecimiento de las áreas de investigación debido a que los frentes de investigación evolucionan constantemente y que, en muchas ocasiones, las delimitaciones entre categorías no están claras.

Tanto en *WoS* como en *Scopus* (y muchas otras bases de datos de datos de ciencia), las revistas pueden asignarse a más de una categoría, mientras que, de hecho, es infrecuente encontrar trabajos que abarquen de alguna manera múltiples temáticas. Y, además, se hace estrictamente necesaria la existencia de categorías multidisciplinares donde dar cabida a las múltiples revistas que existen con un espectro temático muy amplio. De hecho, **Zhang** y **Shen** (2024) señalan que muchas revistas no clasificadas como multidisciplinares quizás deberían clasificarse así debido a la variedad temática que realmente abarcan.

Como referencia, en **Wang** y **Waltman** (2016) se informa de que el número promedio de categorías por revista en *WoS* es de 1,6, y en *Scopus* de 2,1. Pero, si calculamos el promedio de categorías, no por revista, sino por artículo (aplicando la asignación temática a través de las revistas) en el caso de *Scopus* asciende hasta 2,50. Y si analizamos la evolución, se revela un incesante incremento con el avance de los años.

También hay que tener en cuenta que no todos los trabajos que publica una revista serán normalmente de todas y cada una de las categorías a las que está adscrita la revista. De hecho, múltiples estudios señalan la imprecisión en las categorías que las bases de datos atribuyen a las revistas frente a los temas habitualmente tratados en sus artículos. **Thelwall** y **Pinfield** (2024) afirman que esta imprecisión aparece principalmente cuando las revistas no tienen un perfil claramente especializado, por ejemplo, en revistas asignadas total o parcialmente a categorías multidisciplinares o misceláneas.

Estas son algunas de las razones que causan que las clasificaciones basadas en revistas no sean suficientemente precisas, lo cual descalabra, por ejemplo, la aplicación de normalizaciones en los indicadores de citación, arruinando la normalización de las diferencias en los hábitos de publicación y citación entre disciplinas. Se trata de una cuestión crucial para la investigación, y no son pocos los estudios que advierten del problema:

- **Althouse** *et al.* (2009) analizan estas diferencias, cómo han evolucionado a lo largo del tiempo y cómo afectan al cálculo del factor de impacto.
- **Opthof** y **Leydesdorff** (2010) alertan de las consecuencias de ello sobre la evaluación científica de los investigadores.
- **Lancho-Barrantes** *et al.* (2010b) demostraron que los factores de impacto de las grandes bases de datos, que al final son uno de los principales indicadores para evaluar a los investigadores, presentaban una alta correlación con el número de referencias activas (presentes también en la base de datos) por artículo en el periodo correspondiente, evidenciando la necesidad de utilizar medidas de normalización que tuvieran en cuenta esta realidad.
- **Guerrero-Bote** y **Moya-Anegón** (2012) proponen mejorar el indicador SJR añadiendo medidas de normalización adicionales a las que ya incorporaba.
- **Bornmann** y **Leydesdorff** (2017) dan cuenta de la asimetría existente en los cálculos de impacto por citas entre seis grandes disciplinas.
- **Bornmann** *et al.* (2019) comparan varios indicadores bibliométricos sobre los que se aplican distintas normalizaciones.

- **Andersen** (2023) aborda un estudio estadístico que incluye diversos indicadores cienciométricos, identificando algunos riesgos al emplear determinadas técnicas de normalización sobre agregados de disciplinas muy diferentes.
- **Thelwall** y **Pinfield** (2024) resaltan los problemas por la imprecisión en búsquedas y el cálculo de indicadores que se experimentan en los artículos de revistas relacionadas con categorías multidisciplinares o misceláneas.

En todo esto vemos cómo el problema de la normalización se complica cuando tratamos publicaciones en revistas clasificadas como multidisciplinares o que abarcan categorías misceláneas, cuyos artículos, al extender hacia ellos estas clasificaciones por revistas, acabarían sin una temática específica, o, interpretado al contrario, abarcarían una cantidad elevada de temáticas diferentes. Por ello, muchos estudios han centrado sus esfuerzos en intentar clasificar los trabajos publicados en este tipo de revistas (**Glänzel** *et al.* 1999a, 1099b, 2021; **Fang**, 2015; **Zhang** *et al.*, 2022; **Zhang**; **Shen**, 2024).

En la antítesis de la existencia de artículos asignados a muchas categorías existe la doctrina de forzar a que cada artículo se clasifique en una única categoría, lo cual también tiene un alto interés cienciométrico en determinados casos. Por ejemplo, las propuestas de **Milojević** (2020) o **Waltman** y **Van Eck** (2012) centran sus esfuerzos en esto. No obstante son más las propuestas que optan por permitir multiplicidad de asignaciones (por ejemplo, **Fang**, 2015; **Glänzel** *et al.*, 2021; **Zhang** *et al.*, 2022). En concreto, **Zhang** *et al.* (2022) sostienen que es importante permitir asignaciones múltiples porque así lo hacen los propios autores de los trabajos. Y en **Zhang** *et al.* (2016), **Huang** *et al.* (2021) o **Thijs** *et al.* (2021) encontramos argumentos que demuestran la multidisciplinariedad de la ciencia.

Sin embargo, clasificar los artículos en un elevado número de categorías (sobre todo si la relación con algunas de esas categorías es débil), nos devolvería a la imprecisión de la clasificación por revistas, que es, precisamente, lo que se pretende mejorar. En la bibliografía encontramos diversas aproximaciones para limitar el número de categorías:

- **Fang** (2015) propone un umbral de relación, de manera que las categorías con baja relación se descartan.
- **Waltman** *et al.* (2020) limitan el número de posibles asignaciones a las N de mayor grado de relación.
- **Glänzel** *et al.* (2021) desarrollan un criterio más elaborado, donde solo se permiten asignaciones cuyo grado de relación no sea significativamente menor que la asignación previa de mayor grado. Por ejemplo, si la relación con la categoría A es del 90% y con B es del 10%, la categoría B se descarta, pero si con A es del 55% y con B del 45% entonces se admiten ambas.

Añadido a esta idea de limitar el número de categorías, **Glänzel** *et al.* (2021) imponen, además, el criterio de que ninguna de las asignaciones pueda ser a una categoría multidisciplinar general, idea que también se aplica en **Milojević** (2020). A nivel de artículo, es lógico prohibir estas categorías, ya que un trabajo puede abarcar varias temáticas, incluso muchas, pero no todas y cada una de las temáticas de la ciencia.

3. Algoritmos de clasificación a nivel de artículo en la historia

Desde finales del siglo XX, la clasificación de documentos científicos ha sido objeto de múltiples enfoques metodológicos. Uno de los primeros antecedentes lo encontramos en el algoritmo de generación de tesauros para la organización del conocimiento (**Rees-Potter**, 1989). Posteriormente, se exploraron algoritmo basados en la citación entre revistas científicas (**Marshakova-Shaikevich**, 2005; **De-Moya-Anegón** *et al.*, 2006; **Schildt** *et al.*, 2006), así como en la clasificación de documentos específicos como las patentes (**Lai**; **Wu**, 2005) o incluso de revistas completas (**Zhang** *et al.*, 2010).

En paralelo, se desarrollaron algoritmos de agrupamiento automático y detección de comunidades, como los propuestos por **Clauset** *et al.* (2004) y **Blondel** *et al.* (2008), que permitieron identificar estructuras temáticas emergentes en grandes volúmenes de bibliografía científica. Sin embargo, estos enfoques presentan limitaciones importantes: su sensibilidad a la incorporación de nuevos documentos puede alterar significativamente los resultados, y la aleatoriedad inherente a algunos algoritmos puede generar clasificaciones incoherentes incluso con los mismos datos de entrada (**Klavans**; **Boyack**, 2005; 2006; **Waltman**; **Van Eck**, 2012; **Janssens** *et al.*, 2008; 2009).

En este contexto, los algoritmos basados en redes neuronales comenzaron a mostrar un gran potencial para la organización documental. Desde los primeros trabajos (**Guerrero-Bote** *et al.*, 2002), se ha demostrado su capacidad para identificar patrones relevantes tanto en títulos y resúmenes como en textos completos. Estas técnicas han sido aplicadas en entornos de aprendizaje supervisado (**Eykens** *et al.*, 2019) y no supervisado (**Kandimalla** *et al.*, 2021), siendo este último especialmente dependiente de las redes de citación como base para el autoaprendizaje.

A medida que se consolidaban estos enfoques, surgieron propuestas híbridas que combinaban relaciones de citación con análisis textual, con el objetivo de mejorar la precisión de la clasificación temática. Estas metodologías integradas han sido ampliamente estudiadas (**Glenisson** *et al.*, 2005; **Janssens** *et al.*, 2006; 2008; 2009; **Boyack** *et al.*, 2013; **Boyack** & **Klavans**, 2020), y se han beneficiado del desarrollo de técnicas de análisis de contenido (**Boyack** *et al.*, 2011) que incluyen desde el título y las palabras clave hasta el análisis semántico del texto completo.

En cuanto a las relaciones de citación, se han empleado tres mecanismos fundamentales:

- la citación (simple), que vincula una publicación con otra cuando la primera la incluye en su lista de referencias;
- la co-citación, que conecta dos documentos cuando ambos son citados por un tercero; y
- el acoplamiento bibliográfico, que relaciona dos obras que comparten una o más referencias comunes.

Una revisión detallada de estas técnicas puede encontrarse en **Šubelj** *et al.* (2016).

La simplicidad conceptual de la citación (simple) la convierte en una herramienta especialmente útil para gestionar grandes volúmenes de publicaciones con una carga computacional relativamente baja. Esta ventaja ha favorecido su uso en estudios a gran escala, como los realizados por **Boyack** y **Klavans** (2010), **Waltman** y **Van Eck** (2012) y **Klavans** y **Boyack** (2006; 2016). No obstante, no debemos olvidar que algunas publicaciones carecen de citas, y muchas incluyen muy pocas. Hasta un 6,5% de publicaciones de *Scopus* en el año 2020 no tienen ninguna referencia activa, es decir, indexadas en la propia base de datos y, por lo tanto, para las cuales se dispone de alguna información de clasificación; y un 10,4% tienen menos de 3 según se indica en **Álvarez-Llorente** *et al.*, (2024). Esto impide, o, como mínimo, complica, la interpretación de este tipo de relación, al menos sin un esfuerzo adicional de análisis.

No obstante, los algoritmos basados en técnicas híbridas han demostrado ofrecer mejores resultados en términos de precisión y coherencia. **Boyack** y **Klavans** (2020) concluyen que la combinación de relaciones de citación con análisis textual supera tanto a la aplicación aislada de cada técnica como a otras combinaciones basadas exclusivamente en relaciones de citación.

Sin embargo, no todos los estudios coinciden en esta valoración. **Chumachenko** *et al.* (2022) advierten que el análisis textual puede introducir ruido debido a la presencia de frases comunes en múltiples disciplinas. Su propuesta se centra en algoritmos de análisis semántico de textos completos mediante la extracción de conceptos relevantes, utilizando medidas de entropía para distinguir entre conceptos fundamentales y términos genéricos.

En una línea similar, **Sachini** *et al.* (2022) aplican algoritmos de redes neuronales entrenadas con texto para (re)clasificar publicaciones en el ámbito de la inteligencia artificial. Aunque obtienen resultados prometedores, señalan que el alto coste computacional de estas técnicas limita su escalabilidad para conjuntos de datos extensos.

Cabe señalar que algoritmos de agrupamiento y otros algoritmos que generan esquemas de clasificación de forma automática tienden a generar categorías poco coincidentes con las establecidas por las bases de datos, lo que a menudo requiere una intervención manual por parte de expertos. Para facilitar esta tarea, se han implementado programas de visualización como *VOSviewer* (**Van Eck**; **Waltman**, 2010) y *SCImago Graphica* (**Hassan-Montero** *et al.*, 2022), que permiten explorar comunidades temáticas y flujos de información de manera interactiva.

Para mitigar esta variabilidad que introducen los algoritmos de clustering y los basados en redes neuronales —que suelen incorporar factores aleatorios en su lógica— y mejorar la coherencia de las clasificaciones, algunos estudios recientes han optado por utilizar los mismos esquemas temáticos empleados por las grandes bases de datos bibliográficas. Un ejemplo destacado, que ampliaremos en la siguiente sección, es el trabajo de **Milojević** (2020), que asigna categorías temáticas de *WoS* a partir de las referencias citadas en cada artículo, bajo el enfoque de que si la mayoría de las referencias de un documento pertenecen a revistas clasificadas en una categoría determinada, el artículo se asigna a esa misma categoría.

En una línea similar encontramos el algoritmo de clasificación a nivel de documento de **Glänzel** *et al.* (2021) llamado "Modelo paramétrico de múltiples generaciones", que integra muchas de las percepciones que hemos visto, y algunos otros conceptos más o menos innovadores, que también analizaremos en el siguiente apartado.

Recientemente se han publicado dos propuestas de algoritmos de clasificación fuertemente inspirados en la propuesta de **Glänzel** *et al.* (2021), pero incorporando nuevos avances, denominados M3-AWC-0.8 (**Álvarez-Llorente** *et al.*, 2024) y U1-F-0.8 (**Álvarez-Llorente** *et al.*, 2025), los cuales desarrollaremos en mayor detalle en el siguiente apartado.

A modo de recopilación, la figura 1 ordena en una línea temporal las propuestas citadas. Parte de los datos están tomados de **Álvarez-Llorente** (2025). En amarillo se destacan los algoritmos que desarrollaremos en mayor detalle en el siguiente apartado.

Con la proliferación de algoritmos con fundamentos metodológicos diversos y resultados frecuentemente dispares se plantea la necesidad de establecer criterios comparativos sólidos que permitan valorar la calidad relativa de las distintas propuestas. No obstante, **Waltman** *et al.* (2020) sostienen que cualquier sistema de clasificación puede considerarse válido siempre que se adecúe al propósito específico para el que ha sido diseñado, aunque por otro lado enfatizan la necesidad de disponer de métricas objetivas de precisión y de mecanismos que permitan comparar distintos esquemas clasificatorios. Para lograr comparaciones significativas, proponen utilizar una tercera clasificación como referencia, cuya metodología difiera sustancialmente de las otras dos en análisis. Por ejemplo, al contrastar agrupamientos basados en relaciones de citación, sugieren emplear como criterio de evaluación una clasificación fundamentada en análisis textual, y viceversa.

En coherencia con este enfoque comparativo, **Boyack** y **Klavans** (2020) llevaron a cabo un modelo de evaluación que integra tres tipos de métricas de precisión:

- una basada en las referencias contenidas en documentos altamente citados (aquellos que superan las 100 citas),
- otra fundamentada en el análisis textual, y
- una tercera que considera pares de documentos relacionados por vínculos de financiación.

Gracias a esta estrategia, concluyeron que la combinación de relaciones de citación con análisis de contenido textual ofrece resultados superiores, no solo respecto al uso aislado de cada técnica, sino también frente a otras combinaciones que se apoyan exclusivamente en relaciones de citación (aunque, como hemos visto, el coste computacional de estas hibridaciones es muy alto).

Pero, a la hora de evaluar la precisión de una clasificación resulta intuitivo pensar que la idea más acertada sería comparar con una clasificación manual realizada por expertos, o, mejor aún por los mayores expertos sobre cada trabajo, que no son otros que sus propios autores. Este planteamiento se apoya también en la premisa de que una clasificación manual, por su naturaleza radicalmente distinta, puede ofrecer una perspectiva valiosa como punto de comparación.

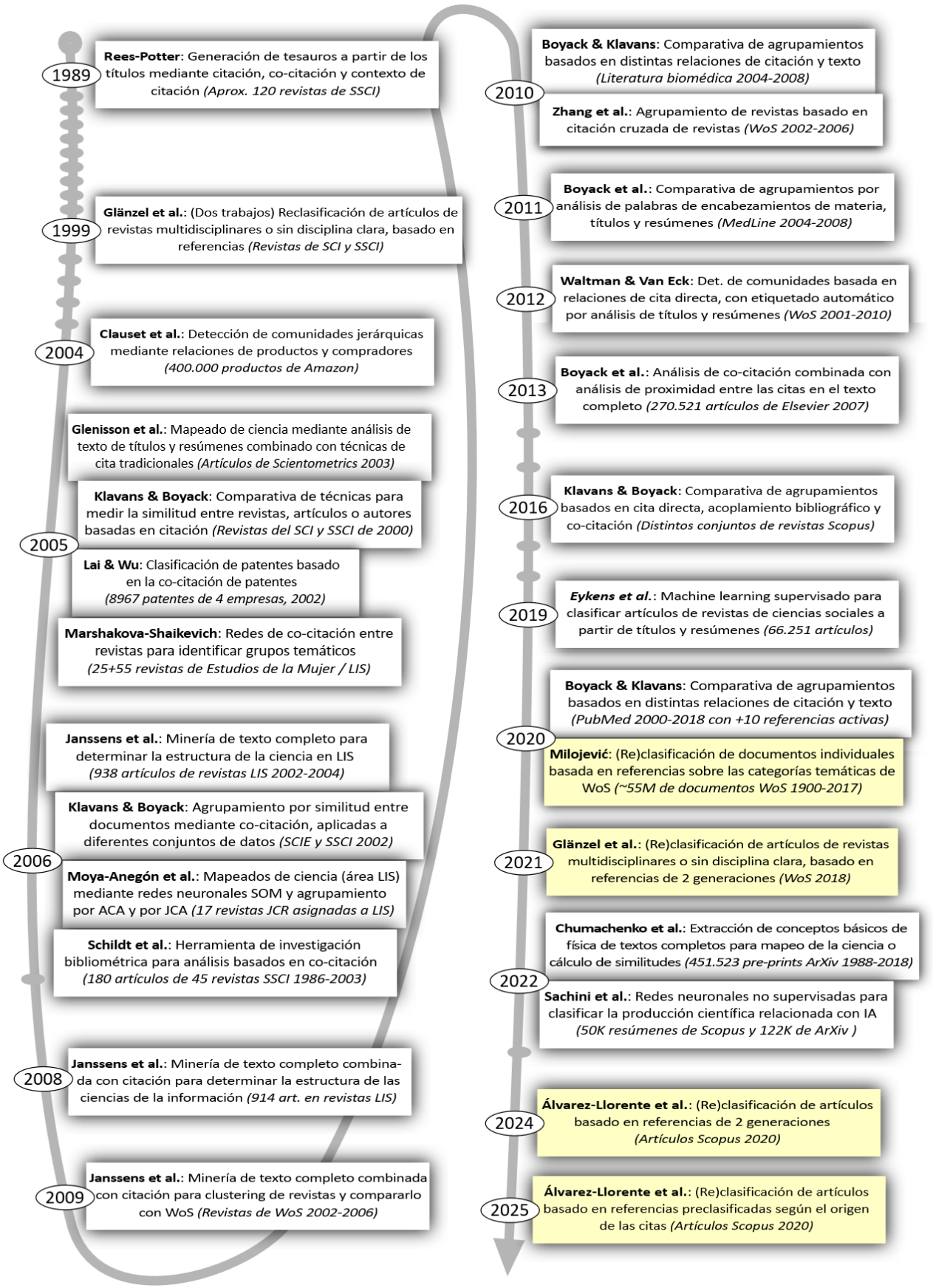

Figura 1. Historia de los algoritmos de clasificación a nivel de documento tratados en este trabajo.

Una de las primeras aproximaciones en esta línea fue la de **Šubelj** *et al.* (2016), quienes recurrieron a la evaluación cualitativa por parte de expertos en cienciometría para valorar la calidad de agrupamientos automáticos, complementando así los indicadores cuantitativos tradicionales.

Otro ejemplo lo encontramos en el estudio de **Milojević** (2020), donde la participación de los autores consistía en seleccionar, para sus propias publicaciones, si consideraban más adecuada la clasificación tradicional por revista según el esquema de *WoS* o la nueva propuesta a nivel de documento desarrollada en el trabajo. La muestra utilizada fue relativamente reducida, compuesta por 142 artículos, todos ellos con asignación temática a una sola categoría y sin carácter multidisciplinar. Además, el estudio incorporó una segunda fase de clasificación manual sobre un conjunto adicional de 100 publicaciones, en la que los autores debían asignar una categoría temática basándose únicamente en el título y el resumen del artículo.

En la misma línea, **Zhang** *et al.* (2022) llevaron a cabo un estudio centrado en publicaciones de la revista *Nature*, en el que compararon tres sistemas de clasificación:

- el etiquetado temático proporcionado por los propios autores al enviar sus manuscritos (basado en la jerarquía temática de la revista, según **McGillivray** y **Astell**, 2019),
- la clasificación temática de *WoS* (a través de *InCites*, basada en las referencias citadas), y
- la clasificación por campos de investigación (*FoR*) de la base de datos *Dimensions*, que emplea técnicas de aprendizaje automático.

Aunque el estudio ofrece una comparación interesante, su principal limitación radica en su aplicabilidad restringida a un único entorno editorial.

Estos trabajos ponen de manifiesto tanto el valor como las limitaciones de las clasificaciones manuales. Por un lado, permiten contrastar los resultados de sistemas automatizados desde una perspectiva humana, posiblemente más cercana a la intención original de los autores. Por otro, presentan problemas evidentes de escalabilidad, ya que su naturaleza manual impide su aplicación a gran escala. Además, el etiquetado por parte de los autores introduce un componente subjetivo que, de hecho, genera incoherencias y discrepancias con respecto a las clasificaciones basadas en referencias o contenido textual (**Shu** *et al.*, 2019; **Zhang** *et al.*, 2022).

A pesar de estas limitaciones, la opinión de los autores se evidencia como un recurso valioso para evaluar la pertinencia de las clasificaciones temáticas. Para que este método resulte realmente útil, sería necesario disponer de un corpus amplio y representativo de publicaciones etiquetadas por sus propios autores, que refleje una diversidad temática, geográfica y disciplinar adecuada.

Con este objetivo, en **Álvarez-Llorente** *et al.* (2023) se presenta la colección *AAC* (*Author's Assignation Collection*), la mayor base de datos hasta la fecha de publicaciones categorizadas por sus autores (autores de correspondencia). Esta colección incluye algo más de 14.000 artículos representativos de distintas áreas del conocimiento y países, clasificados según un esquema denominado "*ASJC* fraccionario", derivado del sistema *ASJC* de *Scopus*, donde, en línea con los principios metodológicos discutidos en el capítulo de contexto, se excluyen el área Multidisciplinar y las categorías misceláneas, y donde permite a los autores asignar su trabajo de forma ponderada a un máximo de cinco categorías. Los resultados del estudio confirman las discrepancias significativas entre las clasificaciones manuales y las generadas automáticamente a partir de referencias o contenido textual. No obstante, la clasificación manual no se plantea como una alternativa operativa para la organización de la ciencia, sino como una herramienta de validación y comparación, ofreciendo un punto de referencia cualitativamente distinto especialmente útil para evaluar la coherencia y precisión de los sistemas automatizados.

AAC (*Author's Assignation Collection*) es la mayor base de datos, hasta la fecha, de publicaciones categorizadas por sus autores (autores de correspondencia)

4. Algoritmos de clasificación a nivel de artículo: el presente

El análisis histórico de las propuestas de algoritmos de clasificación a nivel de artículo ha conducido a la tesis de que una estrategia basada en relaciones de citación, complementada con esquemas de clasificación previamente establecidos, representa una solución equilibrada y viable para abordar la clasificación temática de la producción científica a gran escala. Nos proponemos en este apartado exponer cuatro propuestas recientes de clasificación con estas características, acompañando un breve análisis de sus algoritmos, que quedan resumidos, con sus principales características en la tabla 1.

5. Método práctico para reclasificar artículos de *WoS*

Esta es la propuesta presentada en el artículo titulado "*Practical method to reclassify WoS articles into unique subject categories and broad disciplines*" (**Milojević**, 2020). A partir de aquí nos referiremos a esta propuesta simplemente como la propuesta de Milojević.

Se trata de un algoritmo que permite (re)clasificar a nivel de publicaciones individuales las publicaciones de *WoS*, utilizando, por un lado, la clasificación temática del *WoS* excluyendo las categorías interdisciplinares, y, de forma paralela, a una clasificación en las 14 áreas amplias del *NSF WebCASPAR Broad Field* (**Javitz** *et al.*, 2010), en ambos casos forzando las asignaciones a una única disciplina. El método de reclasificación se basa en las referencias que contienen las publicaciones, siempre y cuando sean "referencias clasificadoras", es decir, que conduzcan a revistas asignadas a una sola categoría y que esta categoría no sea multidisciplinar, de manera que cada publicación se asigna a la categoría a la que pertenezca la mayoría de sus referencias.

Tabla 1. Principales características de los algoritmos de clasificación a nivel de artículo más recientes

Identificación	**Método práctico para reclasificar artículos de *WoS***	**Modelo paramétrico de múltiples generaciones**	***M3-AWC-0.8***	***U1-F-0.8***
Referencia (y año)	**Milojević** (2020)	**Glänzel** *et al.* (2021)	**Álvarez-Llorente** *et al.* (2024)	**Álvarez-Llorente** *et al.* (2025)
Esquema	*NSF WebCASPAR Broad Field // WoS*	Leuven-Budapest modificado	*ASJC* fraccionario	*ASJC* fraccionario
Áreas / Categorías	14 // 252[1]	15 / 73	26 / 285	26 / 285
Máx. categorías asignadas	1	3	5	5
Tipo de relación	Referencias	Referencias de 2 generaciones	Referencias de 2 generaciones	Referencias preclasificadas por sus citadoras
Dataset	*WoS Core Collection* 1900-2017 con alguna referencia activa	*WoS Core Collection* 2018 con alguna referencia activa	*Scopus* 2020 con más de 2 referencias activas	*Scopus* 2020 con más de 2 referencias
Tamaño dataset	45 millones	No especificado (aprox. 3,6M)	3.034.904	3.121.740

Se aplica a la colección completa de publicaciones de *WoS* hasta 2017 que contienen alguna referencia a elementos de esa misma colección, por lo que muchas publicaciones no pueden ser reclasificadas. Sucesivas iteraciones en la aplicación del método permiten incrementar el número de publicaciones reclasificadas.

Entre las principales bondades de este algoritmo destaca su simplicidad, que permite una rápida ejecución sin grandes recursos. Pero el hecho de que asigne categorías únicas desaconseja su uso para propósitos de evaluación de la investigación, como se reconoce en las conclusiones, mientras que defiende su potencial para estudios de bibliometría descriptiva o de ciencia de la ciencia.

El hecho de que la propuesta de Milojević asigne categorías únicas desaconseja su uso para propósitos de evaluación de la investigación

También se admite una pérdida de precisión debida a la limitación de descartar las referencias que no sean clasificadoras, es decir, referencias a revistas asignadas a más de una categoría.

Y no debemos olvidar que solo es aplicable a los artículos que presenten referencias activas.

[1] A diferencia de los otros algoritmos de la tabla, se trata de dos clasificaciones paralelas, una en las 14 áreas amplias del *NSF WebCASPAR* y otra en las 252 categorías de *WoS*, no se trata de una única clasificación con una jerarquía de 14 áreas amplias divididas en 252 categorías menores.

Cabe destacar que en este trabajo, a la hora de evaluar la precisión del resultado, entre otras herramientas, se utilizan pequeñas colecciones de artículos categorizados manualmente por sus autores.

6. Modelo paramétrico de múltiples generaciones

Encontramos esta propuesta en el artículo titulado "*Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research*" (**Glänzel** *et al.*, 2021), aunque no es el único, ni siquiera el principal, propósito de este trabajo. A partir de aquí nos referiremos a esta propuesta simplemente como la propuesta de Glänzel.

Se trata de un algoritmo de (re)clasificación a nivel de documento que, desde nuestro punto de vista, es el que mejor integró en su momento todas las sensibilidades que hemos revisado en los apartados previos, incorporando, además, conceptos innovadores. Utiliza como punto de partida la clasificación por revistas de *WoS*, adaptada al esquema *Leuven-Budapest* modificado (**Glänzel** *et al.*, 2016), que consta de 16 categorías y 74 subcampos, compatibles con el esquema de categorías de *WoS* y de los *JCR*. Aunque de forma teórica propone reclasificar los artículos utilizando relaciones de citación de múltiples generaciones, en la práctica lo hace únicamente con dos (primera y segunda generación: las referencias y las referencias de las referencias), experimentando con 3 posibles ponderaciones sobre ambas generaciones que dan lugar a 3 modelos: M1, solo con las referencias de primera generación; M2, solo con las de segunda generación; y M3, donde ponderan las de primera generación por 0,618 y las de segunda por 0,382[2].

A diferencia de la propuesta de Milojević, admite asignaciones a múltiples categorías, hasta un máximo de 3, y de manera ponderada y normalizada a 1 (la suma de las ponderaciones es 1), pero estableciendo un umbral que rechaza la asignación de una nueva categoría si el peso de esta es inferior a 2/3 del peso de la anterior. La idea es que si una categoría tiene un peso significativamente menor que la anterior asignada, ya no se asigna.

La propuesta de Glänzel admite asignaciones a múltiples categorías, hasta un máximo de 3, y de manera ponderada y normalizada a 1 (la suma de las ponderaciones es 1), pero estableciendo un umbral que rechaza la asignación de una nueva categoría si el peso de esta es inferior a 2/3 del peso de la anterior

También impone la limitación de que no se admite la asignación a la categoría multidisciplinaria X0 *Multidisciplinary sciences* (y consecuentemente tampoco al área amplia *Multidisciplinary sciences*, que únicamente contiene dicha categoría).

Tras los correspondientes análisis de precisión se determina que los mejores resultados se obtienen mediante el modelo M3.

[2] Estos valores responden a los criterios P1+P2=1 y $P1^2$=P2, de manera que la segunda generación tiene una atenuación en su peso que es el cuadrado de la primera.

Entre los aspectos más negativos de esta propuesta podemos apuntar de nuevo la existencia de numerosas publicaciones con muy pocas o incluso ninguna referencia activa, las cuales no se pueden reclasificar. También informa de imprecisión a la hora de reclasificar artículos con alto grado de multidisciplinariedad, para lo cual propone hibridar con métodos de análisis de texto completo.

7. Algoritmo *M3-AWC-0.8*

Esta es la propuesta presentada en el artículo "*New fractional classifications of papers based on two generations of references and on the ASJC Scopus scheme*" (**Álvarez-Llorente** *et al.*, 2024).

Partiendo del algoritmo de Glänzel, con sus tres modelos M1, M2 y M3, propone mejorarlo desde varias perspectivas, ensayando el comportamiento de varios ajustes. En primer lugar se prueba un parámetro denominado "promediado" que trata de atenuar las posibles descompensaciones entre el número de referencias que puede tener un artículo (referencias de primera generación) y el número de referencias (de segunda generación) que pueden tener algunas de estas referencias debido a las diferencias en los hábitos de citación entre disciplinas, y que podrían causar desviaciones temáticas por sobrerrepresentación. A la vez, se examina el comportamiento de las clasificaciones resultantes al emplear diferentes umbrales para admitir asignaciones múltiples, pasando del umbral fijo de 2/3 de la propuesta primitiva de Glänzel a tres posibles umbrales, 1/2, 2/3 y 4/5 (0,8) con los que experimentar. Por último, cada una de estas variaciones se combina con dos métodos de contabilizar la pertenencia de las referencias a las categorías, lo que se conoce como método *Full-Counting* (si una revista está asignada a N categorías, tiene un peso de 1 en cada una de ellas, tal y como se aplica en el algoritmo de Glänzel), y el método *Weighted-Counting* o asignaciones fraccionarias (si una revista está asignada a N categorías, tiene un peso de 1/N en cada una de ellas).

En este caso el esquema de referencia es el *ASJC* fraccionario (**Álvarez-Llorente** *et al.*, 2023), es decir, el *ASJC* de *Scopus* sin el área Multidisciplinar ni categorías misceláneas, admitiéndose hasta 5 asignaciones simultáneas.

El trabajo contiene un concienzudo análisis de todas las variaciones que resultan de la combinación de los parámetros experimentados, centrándose de manera especial en el comportamiento de los artículos de revistas multidisciplinares y de categorías misceláneas, para los cuales se recurre, entre otras herramientas, a la comparación con la clasificación por autores *AAC* (**Álvarez-Llorente** *et al.*, 2023).

Finalmente se determina que los resultados más adecuados se obtienen con el modelo M3 de doble generación, con aplicación de promediado, usando el método *Weighted-Counting* y con el umbral 0,8, combinación que da nombre al algoritmo resultante.

En las conclusiones se resalta que las clasificaciones resultantes del algoritmo muestran mayor homogeneidad que la clasificación por revistas (adaptada al modelo *ASJC* fraccionario), mayor coincidencia con la *AAC*, y otras características cienciométricas deseables.

No obstante se reconoce como una de sus principales limitaciones, una vez más, el no desdeñable número de publicaciones con bajo número de referencias. De hecho, se autoimpone la restricción de no reclasificar aquellos documentos que no cuenten con un mínimo de 3 referencias activas entre las dos generaciones, por considerar que en ellos las referencias no aportan un nivel suficiente de significación.

Las clasificaciones resultantes del algoritmo *M3-AWC-0.8* muestran mayor homogeneidad que la clasificación por revistas (adaptada al modelo *ASJC* fraccionario), mayor coincidencia con la *AAC*, y otras características cienciométricas deseables

8. Algoritmo *U1-F-0.8*

Encontramos esta propuesta en el artículo "*New paper-by-paper classification for Scopus based on references reclassified by the origin of the papers citing them*" (**Álvarez-Llorente** *et al.*, 2025), íntimamente ligado a **Álvarez-Llorente** *et al.* (2024), de manera que el algoritmo *U1-F-0.8* se propone como una alternativa que mejora las limitaciones del *M3-AWC-0.8*, y, al igual que este, utiliza el esquema de clasificación *ASJC* fraccionario asignando ponderadamente un máximo de 5 categorías por artículo. Esto permite una fácil comparativa entre ambos algoritmos y también con la clasificación manual *AAC* realizada por los autores de correspondencia.

U1-F-0.8 es un algoritmo de clasificación por referencias de una sola generación, pero incorpora una idea ya adelantada y esbozada en estudios previos como **Guerrero-Bote** *et al.* (2007), **Lancho-Barrantes** *et al.* (2010a), **Ding** *et al.*, (2018) o por **Li** *et al.* (2019), según la cual, en una relación de citación, no solo ocurre que la publicación citadora tiende a acercarse a la clasificación de la citada, sino que también esta relación tiene el significado contrario, es decir, la publicación citada es atraída temáticamente por la citadora.

Por ello, antes de realizar la (re)clasificación por referencias al estilo clásico, el algoritmo efectúa un paso previo de preclasificación de las referencias, asignándolas a las categorías de los artículos que las han citado. Como en este paso previo las relaciones de citación se interpretan en sentido contrario al tradicional, es posible otorgar una categoría inicial a todas las referencias, tanto activas como no activas, lo cual permite (re)clasificar en la fase posterior un mayor número de publicaciones (sin la limitación de que sus referencias sean activas). No obstante, en este caso se vuelve a aplicar la autolimitación de no reclasificar las publicaciones con menos de 3 referencias.

Una vez efectuada la reclasificación por referencias, la colección de artículos queda reclasificada con una ponderación de pesos en las distintas categorías de manera diferente. Entonces se realiza una segunda iteración del algoritmo, pero previamente se eliminan de la ponderación de pesos resultante los de aquellas categorías que inicialmente no estaban incluidas por la revista. Esta iteración se repite 6 veces[3]. De esta forma, cada artículo recalcula su grado de pertenencia a las diferentes categorías que tenía inicialmente. Finalmente se hace una séptima iteración en la que no se prohíbe la asignación de nuevas categorías.

El fraccionamiento supone un paso de normalización por el que los pesos de las asignaciones al final de cada iteración se dividen entre el número de referencias que tiene cada documento con el objetivo de amortiguar las diferencias en los hábitos de citación entre disciplinas

Los resultados de la sexta y séptima iteraciones dan dos versiones de la clasificación, una con limitación (versión JL, *Journal Limited*) y otra con una iteración sin limitación (versión U1, *Unlimited 1 iteration*), que son evaluadas en el estudio por separado, junto con otros parámetros de ajuste: 3 umbrales de limitación de asignaciones múltiples (los mismos que en el *M3-AWC-0.8*), y dos opciones: con y sin fraccionamiento. El fraccionamiento supone un paso de normalización por el que los pesos de las asignaciones al final de cada iteración se dividen entre el número de referencias que tiene cada documento con el objetivo de amortiguar las diferencias en los hábitos de citación entre disciplinas. El algoritmo se detalla en **Álvarez-Llorente** *et al.* (2025) desde una perspectiva algorítmica formal y desde un punto de vista un poco más matemático. No obstante para facilitar su interpretación, en el Anexo 1 aportamos una interpretación simplificada del mismo.

Los análisis de las variaciones del algoritmo, especialmente centradas en el comportamiento de las publicaciones en revistas multidisciplinares y de categorías misceláneas, determinan que la opción más interesante es la obtenida en la séptima iteración, con fraccionamiento y umbral 0,8, que da nombre al algoritmo final *U1-F-0.8*.

En comparación con el algoritmo *M3-AWC-0.8*, las clasificaciones obtenidas no son muy diferentes, pero *U1-F-0.8*, aporta, además de mayor coincidencia con la *AAC* y otras características cienciométricas deseables, la ventaja de ser aplicable a un mayor número de publicaciones, lo que la proyecta como una interesante propuesta viable para ser adoptada por

En comparación con el algoritmo *M3-AWC-0.8*, las clasificaciones obtenidas no son muy diferentes, pero *U1-F-0.8*, aporta, además de mayor coincidencia con la *AAC* y otras características cienciométricas deseables, la ventaja de ser aplicable a un mayor número de publicaciones

[3] Se repite 6 veces para el conjunto de datos de prueba utilizado en la validación del algoritmo. En realidad la iteración se repite hasta que produce una convergencia en el algoritmo, cuando la diferencia entre la clasificación generada y la obtenida en la iteración anterior es suficientemente pequeña.

las grandes bases de datos para una clasificación por artículos. Podemos encontrar un análisis cienciométrico más profundo de este algoritmo en **Peña-Rocha** *et al.* (2025).

9. Conclusiones

Hemos comenzado este trabajo analizando la problemática que se presenta en las grandes bases de datos como *WoS* o *Scopus* a la hora de categorizar temáticamente los artículos que indexan, ya que lo hacen extendiendo la clasificación de las revistas. Desde la misma aparición de estas bases de datos se han presentado numerosos algoritmos de clasificación individual de artículos con diferentes propósitos.

Los algoritmos de agrupamiento automático y detección de comunidades presentan como principales inconvenientes que generan clasificaciones inestables, con un alto grado de aleatoriedad, basadas en esquemas muy diferentes de los de las bases de datos y que en general requieren de etiquetado manual, por lo que resultan interesantes para objetivos como el estudio de los frentes de investigación, pero no para establecer una categorización estable de las publicaciones.

Como alternativa, los algoritmos que (re)clasifican los artículos por su contenido dentro del mismo esquema temático de las bases de datos se plantean como una opción viable. A la hora de establecer el contenido temático de los artículos, se han propuesto aproximaciones basadas en el análisis de distintos tipos de relaciones de citación y de análisis de texto, siendo estos últimos excesivamente complejos para abordar grandes colecciones. Las relaciones de citación (simple) se proyectan como la opción más viable.

Y así lo encontramos en las propuestas más recientes que hemos examinado. El Método práctico para reclasificar artículos de *WoS* de **Milojević** (2020) tiene como principal limitación que solo admite la asignación a categorías únicas, lo cual resulta insuficientemente preciso. El Modelo paramétrico de múltiples generaciones de **Glänzel** *et al.* (2021) supera esta restricción y constituye una propuesta muy interesante, pero adolece la limitación de que muchos artículos no tienen referencias o tienen muy pocas, lo cual no permite (re)clasificarlo con precisión. Una evolución de este es el algoritmo *M3-AWC-0.8* de **Álvarez-Llorente** *et al.* (2024), que mejora en aspectos de normalización y proporciona clasificaciones más equilibradas, pero mantiene la misma limitación.

Finalmente, el algoritmo *U1-F-0.8* de **Álvarez-Llorente** *et al.* (2025) avanza en la superación de esta restricción reinterpretando el sentido de la relación de citación, con una mayor coincidencia con la *AAC* y características deseables en Cienciometría, constituyendo la mejor propuesta hasta el momento de clasificación de artículos individuales como alternativa a las clasificaciones por revista de las grandes bases de datos.

10. Referencias

Althouse, B. M.; **West, J. D.**; **Bergstrom, C.T.**; **Bergstrom, T.** (2009). Differences in impact factor across fields and over time. *Journal of the Association for Information Science and Technology, 60*(1), 27–34.
https://doi.org/10.1002/asi.20936

Álvarez-Llorente, J. M. (2025). Nuevos algoritmos de clasificación de documentos científicos individuales basados en referencias para mejorar los análisis cienciométricos en las grandes bases de datos de ciencia [Doctoral thesis, University of Extremadura]. Institutional Repository of the University of Extremadura.

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2023). Creating a collection of publications categorized by their research guarantors into the Scopus ASJC scheme. *Profesional de la Información*, 32(7).
https://doi.org/10.3145/epi.2023.dic.04

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2024). New fractional classifications of papers based on two generations of references and on the ASJC Scopus scheme. *Scientometrics, 129*(6), 3493–3515.
https://doi.org/10.1007/s11192-024-05030-2

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2025). New paper-by-paper classification for Scopus based on references reclassified by the origin of the papers citing them. *Journal of Informetrics*, *19*(2), 101647.
https://doi.org/10.1016/j.joi.2025.101647

Andersen, J. P. (2023). Field-level differences in paper and author characteristics across all fields of science in Web of Science, 2000-2020. *Quantitative Science Studies, 4*(2), 394–422.
https://doi.org/10.1162/qss_a_00246

Blondel, V. D.; **Guillaume, J. L.**; **Lambiotte, R.**; **Lefebvre, E.** (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.
https://doi.org/10.1088/1742-5468/2008/10/P10008

Bornmann, L.; **Leydesdorff, L.** (2017). Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *Journal of Informetrics, 11*(1), 164-175.
https://doi.org/10.1016/j.joi.2016.12.001

Bornmann, L.; **Tekles, A.**; **Leydesdorff, L.** (2019). How well does I3 perform for impact measurement compared to other bibliometric indicators? The convergent validity of several (field-normalized) indicators. *Scientometrics, 119*(2), 1187-1205.
http://dx.doi.org/10.1007/s11192-019-03071-6

Boyack, K. W.; **Klavans, R.** (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the Association for Information Science and Technology, 61*(12), 2389–2404. *https://doi.org/10.1002/asi.21419*

Boyack, K. W.; **Klavans, R.** (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies (1)4,* 1570–1585.
https://doi.org/10.1162/qss_a_00085

Boyack, K. W.; **Newman, D.**; **Duhon, R. J.**; **Klavans, R.**; **Patek, M.**; **Biberstine, J. R.**; **Schijvenaars, B.**; **Skupin, A.**; **Ma, N.**; **Börner, K.** (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS One, 6*(3), e18029.
https://doi.org/10.1371/journal.pone.0018029

Boyack, K. W.; **Small, H.**; **Klavans, R.** (2013). Improving the Accuracy of Co-citation Clustering Using Full Text. *J Am Soc Inf Sci Tec, 64*: 1759–1767.
https://doi.org/10.1002/asi.22896

Chumachenko, A.; **Kreminskyi, B.**; **Mosenkis, I.**; **Yakimenko, A.** (2022). Dynamical entropic analysis of scientific concepts. *Journal of Information Science, 48*(4), 561–569.
https://doi.org/10.1177/0165551520972034

Clauset, A.; **Newman, M.**; **Moore, C.** (2004). Finding community structure in very large networks. *Physical Review E, 70*(6).
https://doi.org/10.1103/physreve.70.066111

De-Moya-Anegón, F.; **Herrero-Solana, V.**; **Jiménez-Contreras, E.** (2006). A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research. *Journal of Information Science, 32*(1), 63–77.
https://doi.org/10.1177/0165551506059226

Ding, J.; **Ahlgren, P.**; **Yang, L.**; **Yue, T.** (2018). Disciplinary structures in Nature, Science and PNAS: Journal and country levels. *Scientometrics*, 116(3), 1817–1852. *https://link.springer.com/article/10.1007/s11192-018-2812-9*

Eykens, J.; **Guns, R.**; **Engels, T. C. E.** (2019). Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches. In: *Proceedings of the 17th International Conference on Scientometrics & Informetrics*, Rome (Italy), 2–5 September, 738–743.
https://hdl.handle.net/10067/1630240151162165141

Fang, H. (2015). Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories. *Knowledge Organization, 42*(3), 139–153.
https://doi.org/10.5771/0943-7444-2015-3-139

Glänzel, W.; **Schubert, A.**; **Czerwon, H.** (1999a). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics, 44*(3), 427–439.
https://doi.org/10.1007/bf02458488

Glänzel, W.; **Schubert, A.**; **Schoepflin, U.**; **Czerwon, H.** (1999b). An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis. *Scientometrics, 46*(3), 431–441.
https://doi.org/10.1007/BF02459602

Glänzel, W.; **Thijs, B.**; **Chi, P. S.** (2016). The challenges to expand bibliometric studies from periodical literature to monographic literature with a new data source: the book citation index. *Scientometrics, 109*, 2165–2179.
https://doi.org/10.1007/s11192-016-2046-7

Glänzel, W.; **Thijs, B.**; **Huang, Y.** (2021). Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research. In: W. Glänzel, S. Heeffer, P. S. Chi, R. Rousseau, *Proceedings of the 18th International Conference of the International Society of Scientometrics and Informetrics (ISSI 2021*), Leuven University Press, 453–464.
https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias3394551&context=SearchWebhook&vid=32KUL_KUL:Lirias&search_scope=lirias_profile&tab=LIRIAS&adaptor=SearchWebhook&lang=en

Gläser, J.; **Glänzel, W.**; **Scharnhorst, A.** (2017). Same data—Different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics, 111*(2), 981–998.
https://doi.org/10.1007/s11192-017-2296-z

Glenisson, P.; **Glänzel, W.**; **Janssens, F.**; **De-Moor, B.** (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management, 41*(6), 1548–1572.
https://doi.org/10.1016/j.ipm.2005.03.021

Gómez-Crisóstomo, M. R. (2011). *Study and comparison of the Web of Science and Scopus (1996-2007)* [*Doctoral thesis, University of Extremadura*]. Institutional Repository of the University of Extremadura.

Guerrero-Bote, V. P.; **De-Moya-Anegón, F.** (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of informetrics*, *6*(4), 674-688.
https://doi.org/10.1016/j.joi.2012.07.001

Guerrero-Bote, V.P.; **De-Moya-Anegón, F.**; **Herrero-Solana, V.** (2002). Document organization using Kohonen's algorithm. Information Processing and Management, 38(1), pp. 79-89.
https://doi.org/10.1016/S0306-4573(00)00066-2

Guerrero-Bote, V. P.; **Zapico-Alonso, F.**; **Espinosa-Calvo, M. E.**; **Gómez-Crisóstomo, R.**; **De-Moya-Anegón, F.** (2007). Import-export of knowledge between scientific subject categories: The iceberg hypothesis. *Scientometrics, 71*(3), 423–441. *https://doi.org/10.1007/s11192-007-1682-3*

Hassan-Montero, Y.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2022). SCImago Graphica: a new tool for exploring and visually communicating data. *Profesional de la información, 31*(5), e310502. *https://doi.org/10.3145/epi.2022.sep.02*

Hassan-Montero, Y.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2014). Graphical interface of the SCImago Journal and Country Rank: an interactive approach to accessing bibliometric information. *El profesional de la información*, 23(3). *https://doi.org/10.3145/epi.2014.may.07*

Huang, Y.; **Glänzel, W.**; **Thijs, B.**; **Porter, A. L.**; **Zhang, L.** (2021). *The comparison of various similarity measurement approaches on interdisciplinary indicators* (pp. 1–24). FEB - KU Leuven

Janssens, F.; **Glänzel, W.**; **De-Moor, B.** (2008). A hybrid mapping of information science. *Scientometrics, 75*(3), 607–631. *https://doi.org/10.1007/s11192-007-2002-7*

Janssens, F.; **Leta, J.**; **Glänzel, W.**; **De-Moor, B.** (2006). Towards mapping library and information science. *Information Processing & Management, 42*(6), 1614–1642. *https://doi.org/10.1016/j.ipm.2006.03.025*

Janssens, F.; **Zhang, L.**; **De-Moor, B.**; **Glänzel, W.** (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management, 45*(6), 683–702. *https://doi.org/10.1016/j.ipm.2009.06.003*

Javitz, H.; **Grimes, T.**; **Hill, D.**; **Rapoport, A.**; **Bell, R.**; **Fecso, R.**; **Lehming, R.** (2010). *U.S. Academic Scientific Publishing. Working paper SRS 11-201*. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.

Kandimalla, B.; **Rohatgi, S.**; **Wu, J.**; **Giles, C. L.** (2021). Large scale subject category classification of scholarly papers with deep attentive neural networks. *Frontiers in Research Metrics and Analytics, 5*, 600382. *https://doi.org/10.3389/frma.2020.600382*

Klavans, R.; **Boyack, K. W.** (2005). Identifying a better measure of relatedness for mapping science. *Journal of the Association for Information Science and Technology, 57*(2), 251-263. *https://doi.org/10.1002/asi.20274*

Klavans, R.; **Boyack, K. W.** (2006). Quantitative evaluation of large maps of science. *Scientometrics, 68*, 475–499. *https://doi.org/10.1007/s11192-006-0125-x*

Klavans, R.; **Boyack, K. W.** (2016). Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984–998. *https://doi.org/10.1002/asi.23734*

Lai, K.; **Wu, S.** (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Management, 41*(2), 313–330. *https://doi.org/10.1016/j.ipm.2003.11.004*

Lancho-Barrantes, B. S.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2010a). The iceberg hypothesis revisited. *Scientometrics, 85*(2), 443–461. *https://doi.org/10.1007/s11192-010-0209-5*

Lancho-Barrantes, B. S.; **Guerrero-Bote, V. P.**; **De-Moya Anegón, F.** (2010b). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science, 36*(3), 371-382. *https://doi.org/10.1177/0165551510366077*

Leydesdorff, L.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2010). Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, 61(2), 352-369. *https://doi.org/10.1002/asi.21250*

Leydesdorff, L.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2015). Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of scopus data (1996–2012). *Journal of the Association for Information Science and Technology*, 66(5), 1001-1016. *https://doi.org/10.1002/asi.23243*

Li, K.; **Chen, P.-Y.**; **Fang, Z.** (2019). Disciplinarity of software papers: A preliminary analysis. *Proceedings of the Association for Information Science and Technology (56)*, 706–708. *https://doi.org/10.1002/pra2.143*

Marshakova-Shaikevich, I. (2005). Bibliometric maps of field of science. *Information Processing & Management, 41*(6), 1534–1547. *https://doi.org/10.1016/j.ipm.2005.03.027*

McGillivray, B.; **Astell, M.** (2019). The relationship between usage and citations in an open access mega-journal. *Scientometrics, 121*, 817–838. *https://doi.org/10.1007/s11192-019-03228-3*

Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative science studies, 1*(1), 183-206. *https://doi.org/10.1162/qss_a_00014*

Opthof, T.; **Leydesdorff, L.** (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of informetrics*, 4(3), 423-430.
https://doi.org/10.1016/j.joi.2010.02.003

Peña-Rocha, M.; **Gómez-Crisóstomo, R.**; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2025). Bibliometrics effects of a new paper level classification. *Frontiers in Research Metrics and Analytics, 10*.
https://doi.org/10.3389/frma.2025.1531758

Rees-Potter, L. K. (1989). Dynamic thesaural systems: A bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management, 25*(6), 677–689.
https://doi.org/10.1016/0306-4573(89)90101-5

Sachini, E.; **Sioumalas-Christodoulou, K.**; **Christopoulos, S.**; **Karampekios, N.** (2022) AI for AI: Using AI methods for classifying AI science documents. *Quantitative Science Studies, 3*(4), 1119–1132.
https://doi.org/10.1162/qss_a_00223

Schildt, H.; **Mattsson, J.** (2006). A dense network sub-grouping algorithm for co-citation analysis and its implementation in the software tool Sitkis. *Scientometrics, 67*, 143–163.
https://doi.org/10.1007/s11192-006-0054-8

Shu, F.; **Julien, C.**; **Zhang, L.**; **Qiu, J.**; **Zhang, J.**; **Larivière, V.** (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics, 13*(1), 202–225.
https://doi.org/10.1016/j.joi.2018.12.005

Šubelj, L.; **Van Eck, N. J.**; **Waltman, L.** (2016). Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods. *PLoS one, 11*(4), e0154404.
https://doi.org/10.1371/journal.pone.0154404

Thelwall, M.; **Pinfield, S.** (2024). The accuracy of field classifications for journals in Scopus. *Scientometrics, 129*(2), 1097–1117.
https://doi.org/10.1007/s11192-023-04901-4

Thijs, B.; **Huang, Y.**; **Glänzel, W.** (2021). *Comparing different implementations of similarity for disparity and variety measures in studies on interdisciplinarity*. FEB Research Report MSI_2103, Report No. MSI_2103.
https://lirias.kuleuven.be/retrieve/610314

Van Eck, N.J.; **Waltman, L.** (2010). Software survey: VOSviewer, acomputer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538.
https://doi.org/10.1007/s11192-009-0146-3

Waltman, L.; **Boyack, K. W.**; **Colavizza, G.**; **Van Eck, N. J.** (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies, 1*(2), 691-713.
https://doi.org/10.1162/qss_a_00035

Waltman, L.; **Van Eck, N. J.** (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology, 63*(12), 2378–2392.
https://doi.org/10.1002/asi.22748

Wang, Q.; **Waltman, L.** (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics, 10*(2), 347-364.
https://doi.org/10.1016/j.joi.2016.02.003

Zhang, J.; **Shen Z.** (2024). Analyzing journal category assignment using a paper-level classification system: multidisciplinary sciences journals. *Scientometrics*, *129*, pp. 5963-5978.
https://doi.org/10.1007/s11192-023-04913-0

Zhang, L.; **Janssens, F.**; **Liang, L.**; **Glänzel W.** (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics, 82*, 687–706.
https://doi.org/10.1007/s11192-010-0180-1

Zhang, L.; **Rousseau, R.**; **Glänzel, W.** (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257-1265.
https://doi.org/10.1002/asi.23487

Zhang, L.; **Sun, B.**; **Shu, F.**; **Huang, Y.** (2022). Comparing paper level classifications across different methods and systems: an investigation of Nature publications. *Scientometrics, 127*(12), 7633–7651.
https://doi.org/10.1007/s11192-022-04352-3

Anexo 1. Algoritmo U1-F-0.8