

**ASIMILACION Y DISEMINACIÓN DE LOS ARTÍCULOS DERIVADOS  
DE LAS TESIS DOCTORALES DE MEDICINA EN LA  
LITERATURA CIENTÍFICA**

**Mercedes Echeverría Morrás**

Tesis doctoral presentada en cumplimiento  
de los requisitos de la Universidad de Salamanca  
para el grado de Doctor.

Director: José Antonio Cordón García  
(Universidad de Salamanca)  
Tutor: David Stuart  
(King's College London)

Salamanca, Diciembre 2015

## **Lista de publicaciones**

### **Artículo de revista:**

Echeverría, M., Stuart, D. y Blanke, T. (2015). Medical theses and derivative articles: dissemination of contents and publication patterns. *Scientometrics*, 102(1), 559-586.

### **Ponencia:**

Echeverría, M., Stuart, D. y Blanke, T. (2015). Potential predictability of references in the identification of derivative articles from doctoral theses. In J. Arora et al. (eds), *10<sup>th</sup> International Caliber. 2015. Shimla (India). Innovative Librarianship: Adapting to Digital Realities*. Gandhinagar (India), INFLIBNET Centre. pp. 171-182.

## **Abstract**

El objetivo de esta tesis es investigar las publicaciones derivadas de las tesis doctorales de medicina. La investigación examina el texto completo de una colección de tesis y artículos publicados por los mismos autores con el fin de determinar a través de un análisis de similitud textual las publicaciones procedentes de las tesis doctorales, denominados *artículos derivados*.

El objetivo de esta investigación consiste en identificar los artículos derivados y en proporcionar información sobre sus características de difusión y asimilación en la literatura científica.

El análisis de similitud textual utiliza la organización discursiva de las secciones de los artículos, IMRaD (Introducción, Metodología, Resultados, Discusión) y Referencias. La aplicación informática empleada para detectar automáticamente la similitud textual entre tesis y artículos es el programa anti-plagio Turnitin. Para determinar qué secciones discursivas resultan predictivas para la identificación de los artículos derivados se realizan análisis de inferencia estadística.

Esta tesis desarrolla el estudio desde tres perspectivas: *identificación de los artículos*, basado en el análisis de similitud textual, *modelos de diseminación*, medidos en términos de visibilidad de las revistas y *modelos de asimilación*, medidos en términos de las citas recibidas por los artículos en la literatura científica.

La tesis analiza la predictibilidad potencial de las referencias bibliográficas en la identificación de los artículos derivados. Se comparan los resultados obtenidos del análisis de similitud textual basado en el contenido con el correspondiente de la sección de referencias bibliográficas.

En conjunto, la investigación examina una serie de cuestiones relacionadas con los artículos derivados, tema que, a nuestro entender, ha sido poco investigado en la literatura científica.

Esta tesis, que tiene como objetivo la identificación de los artículos derivados, podría tener aplicaciones potenciales para evaluar las publicaciones de los doctorandos y proporcionar una visión de la producción científica surgida de las tesis doctorales a las universidades.

## Indice

Lista de publicaciones .....	2
Abstract..3	
Abstract (English).....	4
Indice.....	5
Lista de Tablas.....	8
Lista de Figuras.....	10
Lista de abreviaturas.....	11
Agradecimientos.....	12
1. Introducción.....	13
1.1. Las tesis doctorales como fuente de investigación científica.....	14
1.2. Productividad científica de las tesis .....	15
1.3. Planteamiento de la investigación.....	16
1.4. Preguntas de investigación y objetivos.....	17
1.5. Estructura de la tesis.....	18
2. Revisión bibliográfica.....	20
2.1. Definición de trabajos derivados.....	22
2.2. Publicaciones sobre artículos derivados de las tesis de medicina.....	23
2.3. Perspectivas metodológicas para la identificación de artículos derivados .....	25
2.3.1 Similitud textual basada en el contenido y estructura .....	26
2.3.2. Similitud basada en las citas bibliográficas .....	27
2.3.3. Comparación basada en contenido y citas.....	28
2.4. Indicadores bibliométricos de publicación.....	29
2.4.1. Indicador Factor e Impacto.....	29
2.4.2. Indicadores basados en las citas .....	31
2.4.3. Indicadores basados en la colaboración.....	32
2.4.3.1. Colaboración de directores.....	32
2.4.3.2. Colaboración de doctorandos en equipos de investigación.....	33
2.5. Factores extrínsecos relacionados con las publicaciones.....	34
2.5.1. Políticas de doctorado.....	34
2.5.2. Políticas sobre tesis electrónicas.....	35
2.5.3. Políticas de editores de revistas y tesis electrónicas.....	36
3. Metodología .....	38
3.1. Tamaño de la muestra y composición.....	39
3.1.1. Corpus de investigación.....	39
3.1.2. Fuente de los datos .....	40
3.1.3. Registros de la base de datos ERA .....	40
3.2. Productividad científica de los autores.....	41
3.2.1. Bases de datos bibliográficas .....	41
3.2.1.1Búsquedas bibliográficas.....	41
3.2.1.2Análisis bibliométrico.....	42
3.3 Categorización de las unidades de análisis.....	44
3.4 Descripción computacional del análisis de contenido.....	45

3.5	Validación del instrumento de medida.....	47
3.6	Codificación y estructura del dataset .....	48
3.7	Métodos estadísticos.....	51
3.7.1	Análisis clúster.....	51
3.7.2	Matriz de distancia euclídea.....	51
3.7.3	Curva de ROC (Receiver Operator Characteristics).....	51
3.7.4	Diseño del Gold Standard.....	52
3.8	Indicadores bibliométricos de los artículos.....	52
3.8.1	Modelos de diseminación de los artículos.....	52
3.8.1.1	Indicadores de colaboración.....	52
3.8.1.2	Indicadores de visibilidad.....	53
3.8.1.3	Indicadores relacionados con la fecha de publicación.....	56
3.8.2	Modelos de asimilación de los artículos.....	56
3.8.2.1	Citas recibidas por los artículos.....	56
3.8.2.2	Autocitas.....	57
3.8.2.3	Citas institucionales.....	57
3.8.2.4	Citas externas.....	57
4.	Resultados .....	58
4.1.	Descripción de la base de datos.....	59
4.1.1.	Descripción de las tesis.....	59
4.1.2.	Descripción de la producción científica por tipo de documento.....	59
4.1.3.	Distribución de las tesis y artículos por fechas.....	60
4.1.4.	Frecuencia del número de artículos por autor de tesis.....	60
4.2.	Análisis de similitud textual.....	62
4.3.	Análisis estadístico para la identificación de los artículos derivados.....	64
4.3.1.	Gold Standard.....	64
4.3.2.	Área Bajo la Curva (AUC).....	65
4.3.3.	Determinación del punto de corte.....	66
4.4.	Indicadores basados en la sección de Discusión.....	69
4.4.1.	Índice de Similitud Textual.....	69
4.4.2.	Posición del autor de tesis como primer autor de artículo.....	69
4.4.3.	Distribución del crédito de autoría.....	69
4.4.4.	Coautoría de los directores de tesis.....	69
4.4.5.	Número de autores por artículo.....	72
4.4.6.	Intervalo entre la publicación de artículos y lectura tesis.....	72
4.4.7.	Intervalo entre la publicación de la tesis en OA y primera publicación de un artículo online.....	72
4.5.	Indicadores basados en las Referencias Bibliográficas.....	74
4.5.1.	Índice de Similitud Textual.....	74
4.5.2.	Posición del autor de tesis como primer autor de artículos.....	74
4.5.3.	Distribución del crédito de autoría.....	74
4.5.4.	Coautoría de los directores de tesis.....	74
4.5.5.	Número de autores por artículo.....	77
4.5.6.	Intervalo entre la publicación de artículos y publicación de tesis.....	77
4.5.7.	Intervalo entre la publicación de la tesis en OA y primera publicación de un artículo online.....	77
4.6.	Comparación entre indicadores basados en la sección de Discusión y Referencias .....	79
5.	Análisis bibliométrico .....	80
5.1.	Indicadores relacionados con la visibilidad de las revistas.....	81
5.1.1.	Factor de Impacto. Cuartiles .....	82
5.1.2.	Posición Decil (P10) .....	84
5.1.3.	Factor de Impacto Normalizado (FIN).....	85

5.1.4.	Eigenfactor .....	87
5.1.5.	Article Influence.....	89
5.1.6.	Comparación empírica entre métricas.....	90
5.2.	Indicadores relacionados con el impacto de los artículos.....	91
5.2.1.	Citas recibidas.....	91
5.2.2.	Autocitas.....	94
5.2.3.	Citas recibidas de autores de la Universidad de Edimburgo.....	95
5.2.4.	Citas externas .....	96
5.2.5.	Resumen de las citas recibidas.....	97
5.3.	Resumen análisis bibliométrico.....	98
6.	Discusión.....	100
6.1	Modelos de diseminación de los artículos derivados.....	101
6.1.1.	Indicadores de colaboración.....	101
6.1.1.1	Coautoría y número de autores.....	102
6.1.1.2	Posición del autor de la tesis en los artículos .....	102
6.1.1.3	Distribución del crédito de autoría.....	104
6.1.1.4	Colaboración de los directores como coautores.....	105
6.1.2.	Indicadores de visibilidad.....	107
6.1.2.1.	Factor de Impacto (cuartiles / deciles).....	107
6.1.2.2.	Factor de Impacto Normalizado (FIN).....	108
6.1.2.3.	Eigenfactor.....	109
6.1.2.4.	Article Influence.....	110
6.1.3.	Indicadores relacionados con la publicación de los artículos.....	110
6.1.3.1.	Intervalo entre lectura tesis publicación artículos.....	110
6.1.3.2.	Intervalo entre la publicación de la tesis en OA y primera publicación de un artículo online .....	111
6.2.	Modelos de asimilación de los artículos derivados.....	113
6.2.1.	Citas recibidas por los artículos.....	113
6.2.2.	Autocitas.....	116
6.2.3.	Citas institucionales.....	116
6.2.4.	Citas externas .....	117
6.2.5.	Resumen de las citas recibidas.....	118
6.3.	Predictibilidad potencial de las referencias en la identificación de los artículos derivados.....	119
7.	Conclusiones.....	121
7.	Conclusions (English).....	128
8.	Bibliografía .....	134

Apéndice1. Comparación de listas de artículos derivados basados en la Discusión y Referencias.

Apéndice 2. Publicaciones

## **Agradecimientos**

Esta tesis no hubiera sido posible sin el apoyo de varias personas.

En primer lugar, me gustaría agradecer a David Stuart y José Antonio Cordón por su apoyo y enseñanza. Me gustaría dar las gracias a David Stuart por su dedicación, ayuda y el tiempo destinado a enseñarme a investigar, a José Antonio Cordón por su constante apoyo en la tesis, guía y estímulo durante estos cuatro años, a Tobias Blanke por darme la oportunidad de investigar en el Centre for e-Research (King's College London), por sus valiosas ideas y estímulo. Me gustaría también expresar mi gratitud Mark Hedges por la hospitalidad recibida en el Centro. Me gustaría también agradecer a Ángel Rodríguez su ayuda con los números.

Gracias a mis padres por todo lo que han hecho por mí, hermanos sobrinos, amigos y colegas por su generosa e indefectible ayuda.

## **1. INTRODUCCIÓN**

*“Theses are the most invisible form  
of useful literature and the most  
useful form of invisible literature“*  
[Peter Suber]

## INTRODUCCIÓN

Las tesis doctorales constituyen un componente esencial de la educación de postgrado y una fuente importante de publicaciones científicas para las universidades. Como definición, una tesis es un *trabajo original de investigación* "en el que un autor presenta los hallazgos científicos, en apoyo a la candidatura para el título o cualificación profesional de doctor" (ISO, 1986, p. 679). Desde una perspectiva científica, una tesis es una investigación original que contribuye al avance del conocimiento en una disciplina científico (Quality Assurance Agency for Higher Education, 2011; Real Decreto, 1393/2011). Desde una perspectiva académica, una tesis es un documento de investigación sujeto a examen y necesario obtener el título de doctor (Paillassard, Schöpfel y Stock, 2007).

El desarrollo de los estudios de doctorado en Europa tiene su base en la *Declaración de Bolonia* (Bologna Declaration, 1999), con la creación del Espacio Europeo de Educación Superior (EEES) y la *Agenda de Lisboa* (European Council, 2000), cuyos fundamentos epistemológicos fueron fortalecer la investigación y competitividad de Europa. En este contexto, fue significativa la declaración formulada por la *Conference of Ministers responsible for Higher Education* in Berlin (Berlin Communiqué, 2003) sobre la importancia de la investigación como parte integral de la educación en Europa, con el enfoque de dos ciclos principales en la educación superior, el nivel de doctorado y el tercer ciclo.

El *Seminario de Salzburgo* (European University Association, 2005a) implementó los principios de doctorado basados en la investigación original como componente principal, la importancia de las estrategias institucionales y el tratamiento de los doctorandos como investigadores en fase de formación. Estos puntos fueron incluidos en el *Communiqué of the Conference of European Ministers Responsible for Higher Education*, en Bergen (Bergen Communiqué 2005).

En este contexto, las universidades europeas han experimentado un cambio significativo en el número de estudiantes que terminan el doctorado, con un crecimiento del 43% entre los años (2004-2012) (Eurostat, 2015). Los cambios también se han centrado en una rápida expansión de las escuelas de doctorado europeas, pasando de ser un 29% en 2007 a un 49% en 2010, cuyas principales actividades se han centrado en promover la cooperación entre disciplinas científicas, mejorar las oportunidades de colaboración a nivel internacional e interinstitucional y proporcionar vínculos con la industria y servicios públicos (Sursock, Smidt y Davies 2010).

Quince años después de la *Declaración de Bolonia* (Bologna Declaration, 1999) las estructuras de doctorado se han establecido en gran medida sobre los principios de investigación original, avance en las fronteras del conocimiento, transferencia y intercambio de conocimientos, difusión de la investigación a través de las tecnologías de la información y un nuevo estatus de los doctorandos como investigadores en fase de formación.

### 1.1. Las tesis doctorales como fuente de información científica

Las tesis doctorales han sido clasificadas como *literatura gris*, debido a que se mantienen generalmente fuera de los circuitos comerciales y no han sido sometidas a un proceso formal de revisión por pares. Estos factores, unidos a su tradicional restricción de consulta en las bibliotecas (Andrew, 2004; Copeland, Penman y Milne, 2005; Macduff, 2009) y a su bajo nivel

de uso han sido algunas de sus características, y cuya alternativa de acceso ha sido, con frecuencia, la publicación de partes seleccionadas de las tesis en revistas científicas (Salmi, Gana y Mouillet de 2001, Larivière, 2011; Christianson, Elliot y Massey 2015).

La publicación electrónica de las tesis doctorales ha cambiado significativamente esta situación ofreciendo la posibilidad de acceder al texto completo. El punto de partida de las tesis electrónicas (ETDs) fue cuando las universidades empezaron a exigir que las tesis se depositaran en los repositorios institucionales. La primera institución en requerir su depósito fue *Virginia Polytechnic Institute and State University (Virginia Tech)* en 1997. En la actualidad, un gran número de instituciones de todo el mundo, casi 1.700 (55%) de acuerdo con OpenDOAR (2016) disponen de repositorios donde se depositan las ETDs. La publicación de las ETDs en la web beneficia tanto a sus autores porque mejora la visibilidad de sus investigaciones como a los investigadores porque mejora la accesibilidad mediante una mayor exposición a través de los repositorios institucionales, Google Scholar y herramientas del descubrimiento. Sin embargo, hay una relativa escasez de información respecto a la influencia del OA sobre las tesis doctorales, a pesar de la opinión general sobre la importancia de estos materiales en la investigación académica.

En el actual contexto digital, donde el OA (Open Access) ha mejorado la accesibilidad y el aumento de consultas a las tesis doctorales a texto completo la pregunta es ¿cuál es la razón por la que los investigadores siguen estando motivados para publicar los resultados de sus tesis en revistas científicas?.

La respuesta a esta pregunta está relacionada con el reconocimiento generalizado atribuido *al sistema de revisión por pares* de las revistas científicas, que garantiza la calidad y visibilidad de los resultados científicos. En este sentido, el papel desempeñado por las publicaciones presentadas para la obtención del doctorado podría ser considerado como un indicador de aceptación científica en la defensa una la tesis, posibilitando su evaluación (Salmi, Gana y Mouillet, 2001; Nieminen et al 2007; Christianson, Elliot y Massey, 2015). Además, se observa una creciente tendencia a la difusión de los resultados de la investigación doctoral en revistas científicas antes de la defensa de las tesis (Powell, 2004; Watts, 2012; Christiansen, Elliott y Massey, 2015), lo que sugiere que el proceso de revisión por pares se mantiene como un factor central de la comunicación académica en el entorno electrónico actual.

Además, el creciente aumento de publicaciones antes de la defensa de la tesis revelaría la presión que los estudiantes de doctorado reciben de las instituciones, directa o indirectamente, respecto a la publicación de sus resultados científicos y el papel que desempeñan las publicaciones en la evaluación doctoral. Desde un punto de vista práctico, la publicación de artículos antes de la defensa de una tesis sería percibida positivamente por los candidatos a doctor por los beneficios derivados del sistema de revisión por pares.

Una consecuencia directa de la influencia de las publicaciones de artículos revisados por pares sobre las tesis es el cambio de la estructura de las tesis científicas. El estilo de tesis monográfica está cambiando hacia un nuevo formato donde las publicaciones se conviertan en parte integral de su presentación.

Por otra parte, respecto a la ventaja específica derivada del OA en el aumento de las citas de las tesis, Jones y Andrew expresaron su perspectiva de un posible aumento de las citas, pero sin proporcionar datos estadísticos concluyentes "los datos de los repositorios sugieren un aumento importante del uso y citación de tesis doctorales en la actividad investigadora actual" (2005, p. 198), Copeland, Penman y Milne señalaron que "un mayor uso de las tesis es probable que conduzca a un aumento de las citas, y, consecuentemente mejore el perfil de la investigación" (2005, p. 196), Jones y Andrew, por su parte, examinaron las estadísticas de visitas de las tesis de la Universidad de Edimburgo, indicando que "los datos de los repositorios

sobre las ETDs sugieren un aumento de uso y citación importante de las tesis doctorales, lo que se relaciona con una mejora de acceso y el descubrimiento de recursos científicos atribuido a los sustitutos digitales "(2005, p.198).

En 2007 un estudio longitudinal realizado por Larivière, Zuccala y Archambault "*The declining scientific impact of theses: implications for electronic thesis and dissertation repositories and graduate studies*" demostró que "la cifra absoluta de citas a las tesis doctorales comenzó a descender en 2001, y que la caída del número de citas de las tesis comenzó a acelerarse a partir de 2001. Esto demostraría que los repositorios electrónicos, en el mejor de los casos, no han hecho más que ralentizar la tendencia a una disminución de las citas de las tesis "(2007, p. 115).

En general, aunque el OA ha mejorado la visibilidad, accesibilidad y comunicación de las tesis doctorales, sin embargo parecería que este tipo de publicaciones no han sido incluidas en el sistema de recompensas y citas por la comunidad científica, existiendo una tendencia creciente entre los estudiantes de doctorado a difundir los resultados de su investigación a través de las revistas científicas, debido a las ventajas percibidas del sistema de revisión de pares.

## **1.2. Productividad científica de las tesis**

El contexto de la educación doctoral está caracterizado por un crecimiento tanto de la diversidad de programas como por el número de estudiantes de doctorado.

El doctorado convencional a tiempo completo ya no es la única vía para alcanzar un título de doctor, hay diferentes formas de producción y diferentes 'rutas' para la obtención del doctorado (Park, 2007; Sharmini et al., 2015). La educación doctoral se enfrenta a cambios en el mercados de trabajo con programas que integran doctorados profesionales, cuyo objetivo es promover el conocimiento dentro de un contexto práctico (Hoddell, Street y Wildblood 2002). Otro modelo para alcanzar el título de doctor son las 'tesis por compendio de publicaciones' que consisten entera o predominantemente en publicaciones de artículos de revistas o libros, revisados por pares, cuyo contenido está relacionado con el proyecto de tesis regulado en los programas de doctorado (Powell, 2004; Davies y Rolfe, 2009; Kamler, 2008). Para optar a esta modalidad, las tesis españolas deberán cumplir los requisitos establecidos por el Real Decreto 1393/2007 sobre proyecto de tesis doctoral y con el plan de investigación que consta en el documento de actividades del doctorado de los programas regulados por el Real Decreto 99/2011.

Cualquiera que sea la forma de doctorado adoptada, el principal requisito es realizar una investigación original que contribuya al conocimiento científico. Esto conlleva la consecuencia implícita de que lo que merece la pena ser investigado y explorado merece la pena ser difundido y publicado. Como resultado, las universidades están animando a los estudiantes de doctorado a publicar durante o como parte de su doctorado (Kamler, 2008; Dhaliwal, Singh y Bhatia, 2010; Sharmini et al., 2015). En el Reino Unido, aunque no hay una normativa por la que se requieren publicaciones científicas para obtener el título de doctor, se observan presiones sobre los doctorados para publicar en las disciplinas científicas. (Christianson, Elliot y Massey, 2015). En general, se espera que un candidato haya publicado algún artículo antes de la lectura de su tesis.

Esta práctica de publicar durante el doctorado se podría argumentar desde tres perspectivas, los intereses y responsabilidades comunes de profesores y estudiantes respecto al desarrollo de la investigación mediante la publicación, la presión que reciben los estudiantes para participar en proyectos de investigación y la competitividad del mercado de trabajo. Es probable que la presión por publicar que reciben los profesores, se transfiera también a su participación en las publicaciones con doctorandos, como una forma de aumentar su productividad investigadora.

A pesar de la importancia de las publicaciones generadas en el periodo de doctorado, la atención que han recibido en la literatura ha sido relativamente escasa. La mayoría de los estudios publicados se han llevado a cabo a nivel micro, lo que proporciona una visión parcial sobre las características de los artículos relacionados con las tesis doctorales en términos de productividad, identificación e indicadores bibliográficos. De hecho, existe una escasez de datos empíricos respecto a la identificación y características de los artículos derivados de tesis doctorales, siendo todavía como un tema de investigación inexplorado.

Esta tesis investiga los artículos científicos derivados de las tesis de medicina cuyo formato corresponde a una monografía. El objetivo es proporcionar datos empíricos sobre la identificación de los artículos derivados a través de un análisis de similitud textual del texto completo de las tesis y de los artículos con los que comparten autoría. Otros objetivos de esta investigación son analizar los patrones de difusión y asimilación de los artículos derivados en la literatura científica.

### **1.3. Planteamiento de la investigación**

El propósito de esta investigación no es el estudio de las tesis en sí mismas, sino la investigación de las publicaciones procedentes o derivadas de las mismas, lo que implica que la tesis es la fuente de información y publicaciones son el resultado del proceso de investigación doctoral.

El punto de partida es la tesis cuyo formato en una monografía, basada en gran medida en un proyecto de investigación supervisado. El término 'tesis monográfica' se refiere también a aquellas tesis que pueden incluir publicaciones adjuntas. Esta investigación se centra en las tesis doctorales de medicina con una alta productividad de publicaciones de los estudiantes de doctorado [Véase Metodología 3.2], lo que facilita el espectro de análisis de los *artículos derivados*.

El término '*artículos derivados*', aunque su significado pueda ser ampliamente entendido, los parámetros bajo los que una publicación se convierte en una obra derivada a menudo no están claramente definidos. El término '*artículos derivados*' específicamente se emplea en esta tesis para referirse a los artículos 'basados en o extraídos de una tesis'. Explícitamente, Dhaliwal, Singh y Bhatia (2010) consideran que un artículo es '*derivado de una tesis*' cuando satisface ambos de los siguientes criterios (i) al menos, "uno de los autores de la tesis" (sic) es autor de la publicación, (ii) al menos, alguno de los resultados de la tesis es a su vez un resultado de la publicación.

En esta investigación, los artículos derivados se definen en función de tres factores que deben producirse conjuntamente. Un artículo científico será derivado de una tesis si se publica por el mismo autor de la tesis, comparte similitud textual y de contenido con la tesis y se produce durante el periodo de doctorado o inmediatamente después de la finalización de la tesis.

Esta tesis investiga una colección de tesis de medicina para determinar la identificación de artículos derivados a través de un análisis de similitud textual basado en el texto completo de las tesis y de los artículos de los mismos autores. Un objetivo de esta investigación es proporcionar una visión de los modelos de reutilización textual de las tesis en las posteriores publicaciones de artículos científicos. Una publicación de este tipo se considera a menudo el paso lógico en el proceso de investigación, ya que tanto maximiza el impacto de un trabajo de investigación y proporciona un medio al investigador demostrar su impacto.

#### 1.4. Preguntas de investigación y objetivos

El objetivo de esta tesis fue determinar la *identificación de artículos derivados* de las tesis de medicina y proporcionar información sobre las características de los *modelos de difusión y asimilación* de sus contenidos en la literatura científica.

Tres preguntas de investigación generales se exploran esta tesis, a cada una de las cuales les corresponde diferentes objetivos de la investigación:

a) *¿Cuál es el umbral y distribución de similitud textual entre tesis y artículos del mismo autor que determinan la identificación de los artículos derivados?*

Los objetivos fueron:

- Explorar nuevos enfoques metodológicos de minería textual, basados en técnicas estadísticas y computacionales, destinados a medir la distribución de similitud textual entre tesis y artículos con el fin de identificar los artículos derivados.
- Examinar los niveles de similitud textual de los artículos, mediante un análisis de clúster de acuerdo con las diferentes secciones discursivas, Introducción, Metodología, Resultados, Discusión (IMRaD). Este análisis permitirá saber qué secciones argumentativas producen mayor densidad de similitud textual y detectar, de esta forma, las secciones que podrían determinar la identificación de artículos derivados.
- Evaluar la predictibilidad potencial de referencias bibliográficas comunes entre tesis y artículos, utilizando el análisis de similitud textual, con el fin de medir la capacidad de referencias como factor de detección de artículos derivados.

Preguntas exploradas en esta investigación relacionadas con estos objetivos fueron:

- ¿Cuántos de los artículos de investigación doctoral son artículos derivados?
- ¿Qué índices de similitud textual pueden predecir si un artículo es derivado?
- ¿Qué sección/es son predictivas para la identificación de artículos derivados?

b) *¿Cuáles son las características de los modelos de diseminación de los contenidos de las tesis doctorales en la literatura científica?*

Los objetivos fueron:

- Examinar críticamente los niveles de visibilidad de las revistas de los artículos derivados, basados en la influencia de indicadores como el Factor de Impacto y posicionamiento de las revistas.
- Examinar críticamente la influencia de las revistas de los artículos derivados.
- Analizar los indicadores de colaboración en los artículos derivados.
- Analizar los indicadores relacionados con los modelos de publicación de los artículos derivados.

Preguntas exploradas sobre los artículos derivados relacionados con estos objetivos fueron:

- ¿Cuáles son los indicadores de visibilidad y posicionamiento de las revistas?
- ¿Cuáles son los indicadores Eigenfactor y Article Influence de las revistas de estos artículos?
- ¿Cuáles son los indicadores de coautoría?
- ¿Cuál es el orden de los autores en la firma de los artículos derivados?
- ¿Cuál es la distribución del crédito de autoría?
- ¿Cuál es la participación de los directores como coautores?
- ¿Cuál es el intervalo temporal entre la publicación de los artículos derivados y la lectura de las tesis?
- ¿Cuál es el intervalo entre la publicación de la tesis en OA y la primera publicación de los artículos derivados en OA?

c) *¿Cuáles son las características de los modelos de asimilación de los contenidos de las tesis doctorales en la literatura científica?*

- Examinar críticamente los niveles de impacto de los artículos derivados, medidos en términos de citas recibidas procedentes de la comunidad científica.
- Comparar y evaluar las citas recibidas por los artículos derivados y no-derivados del mismo autor.

Preguntas exploradas en esta investigación en relación con estos objetivos fueron:

- ¿En qué medida los artículos derivados reciben citas?
- ¿De quién proceden las citas recibidas? ¿Existe relación entre el autor de la tesis y la persona que cita el artículo derivado?
- ¿Cuáles son los niveles de autocitas, citas institucionales y citas externas tanto de los artículos derivados como de los no-derivados?

## **1.5. Estructura de la tesis**

Esta tesis se compone de cinco secciones principales:

- Revisión bibliográfica
- Metodología
- Resultados
- Discusión
- Conclusiones

### **Revisión bibliográfica**

La revisión bibliográfica introduce el concepto de artículos derivados y examina los enfoques metodológicos para la identificación de estos artículos. Define el marco y los factores que influyen en la aparición de estas publicaciones y presenta críticamente los estudios anteriores sobre artículos derivados de las tesis de medicina.

## **Metodología**

Esta sección contiene dos partes: una descripción del diseño, procedimientos analíticos y métodos estadísticos utilizados para la identificación de estos artículos. Una segunda parte describe los indicadores bibliométricos utilizados para analizar los modelos de difusión y asimilación de estos artículos.

## **Resultados**

Este capítulo presenta los datos obtenidos de la aplicación de diferentes análisis estadísticos. Se presentan los datos con respecto a la identificación, las características de los modelos de publicación y difusión de estos artículos, medidos por indicadores de visibilidad, colaboración y publicación, y finalmente las características de su asimilación en la literatura científica, medidos por indicadores basados en las citas recibidas.

## **Discusión**

La sección de Discusión describe los resultados obtenidos, analizándolos en relación con las diferentes aportaciones y estudios de bibliometría.

## **Conclusiones**

Este último capítulo resume las principales aportaciones de la tesis, expone algunas limitaciones del enfoque utilizado, y propone posibles vías de trabajo en el futuro, así como las aplicaciones que podría generar esta investigación.

## **2. REVISIÓN BIBLIOGRÁFICA**

*Caminante no hay camino,  
se hace camino al andar.  
[Antonio Machado, 1875-1939]*

## REVISIÓN BIBLIOGRÁFICA

Esta sección examina la literatura relevante sobre los artículos derivados de las tesis de medicina desde varias perspectivas: concepto, enfoques metodológicos para su identificación, indicadores bibliográficos y factores que influyen en su publicación. Esta revisión presenta una síntesis crítica de los trabajos previos sobre los temas abordados en esta investigación.

La materia de la investigación los *artículos derivados de las tesis doctorales* se ha explorado en bases de datos, revistas especializadas e Internet, pero se ha observado que existen pocas publicaciones que aborden este tema. La razón principal sería su relativa novedad como objeto de investigación y el enfoque del análisis del textual de tesis y artículos. La ausencia de estudios sobre estas publicaciones constituye por tanto una limitación y una oportunidad para avanzar en la comprensión del papel de estas publicaciones académicas.

La revisión de la literatura está dividida en cuatro secciones:

La primera sección examina:

- El concepto de *artículos derivados*
- *Los artículos derivados* de las tesis de medicina en la literatura científica

La segunda sección examina:

- Los enfoques metodológicos para la identificación de artículos derivados basados en:
  - Contenido y estructura
  - Referencias
  - Comparación entre los métodos basados en contenido y referencias

La tercera sección examina:

- Los indicadores bibliográficos basados en
  - El Factor de Impacto (FI)
  - Citas recibidas
  - Colaboración de autores

La cuarta sección examina:

- Los factores que influyen en la publicación de artículos derivados:
  - Políticas de doctorado
  - Políticas de OA relacionadas con las tesis electrónicas
  - Políticas de los editores de revistas científicas

## 2.1. Definición de los trabajos derivados

El primer paso consiste en determinar el concepto de '*artículo derivado*' de tesis doctorales. Una dificultad encontrada en la literatura ha sido la falta de una terminología consistente sobre esta materia. Este hecho podría estar relacionado tanto con la ambigüedad del término '*derivado*', como los límites tradicionales de las tesis doctorales impresas, categorizadas como *literatura gris*. Hoy en día, las tesis digitales y el análisis de textual asistido por ordenador ofrecen la posibilidad de difuminar estos límites y poder abrir procesos de examen que no se podrían haber llevado a cabo con las publicaciones impresas, proporcionando información sobre las características de este tipo de documentos.

El término '*trabajos derivados*' en la literatura está asociado a diferentes conceptos:

### a) Trabajos derivados y OA.

*La Declaración de Bethesda sobre Publicación en Acceso Abierto* (Bethesda Statement, 2003) reforzó el énfasis de la difusión sin barreras de obras científicas en OA. Entre los principios esbozados se detalla la reutilización de los trabajos científico-académicos y la elaboración y distribución de *obras derivadas*.

*Creative Commons* (2008) proporciona una definición sobre '*trabajos derivados*' en el Título 17 "Un trabajo basado en el trabajo o en la palabra de otras obras preexistentes, tales como una traducción, arreglo musical, dramatización, versión de imágenes movimiento, grabación de sonido, reproducción de arte, resumen, o cualquier otra forma en la que una obra puede ser reformulada, transformada o adaptada".

Respecto a la implementación en el Reino Unido del Título 17 de *Creative Commons*, la posición adoptada ha sido considerar que "[una licencia de obra derivada] da al usuario los derechos para copiar, distribuir y mostrar las obras, y hacer *trabajos derivados*, siempre y cuando el autor original conceda la correspondiente autorización, y el trabajo se utilice para fines no comerciales" (Andrew, 2004, p. 13).

Según el UK Copyright Service (2012) '*obra derivada*' es una nueva obra basada en (derivado de) otra obra. Legalmente sólo el propietario del derecho de autor tiene el derecho de autorizar las adaptaciones y reproducciones de su trabajo. A menos uno sea el propietario de la obra original, es probable que necesite el permiso del propietario de los derechos de autor antes de realizar una *obra derivada* ".

En sentido estricto, este tipo de '*trabajos derivados*' cuyas palabras clave son "transformación o adaptación", implican bastante creatividad y originalidad. De acuerdo con la legislación británica sólo el propietario de los derechos de autor tiene la potestad de autorizar las adaptaciones y reproducciones de su trabajo, lo que incluye también la realización de un trabajos derivados.

### b) Artículos derivados y editores de revistas

El concepto de artículos derivados de tesis doctorales es bien conocido por los editores de revistas. La publicación de un artículo derivado de una tesis es una práctica habitual, cuya aceptación o rechazo podría variar en función de las políticas aplicadas por las revistas respecto a lo que constituye "*prior publication*" (ver sección 2.5.3.).

Algunos editores como, *American Psychological Association*<sup>1</sup>, *DKS Publications*<sup>2</sup>, *MAT Journals*<sup>3</sup> proporcionan directrices para la publicación de artículos cuyo contenido está basado en una tesis doctoral. En este sentido, las políticas de estos editores van dirigidas a reconocer la investigación original de las obras derivadas de tesis depositadas en OA en los repositorios.

c) Artículos derivados y universidades

Las publicaciones basadas en tesis doctorales son una de las fuentes de publicación de las universidades (Kamler, 2008). En este sentido, las universidades son conscientes de la importancia estratégica de la difusión de la información sobre las políticas aplicadas por los editores en relación con la publicación de artículos procedentes de las tesis doctorales. Bibliotecas universitarias como MIT Libraries<sup>4</sup> Leicester Library<sup>5</sup> o Library of University of Michigan<sup>6</sup>.

Paralelamente las universidades están exigiendo cada vez más en sus programas de doctorado que los estudiantes publiquen los resultados de sus tesis (ver sección 2.5.1). Algunos programas de doctorado son sometidos a procesos de evaluación, medidos entre otros criterios, por los artículos publicados durante la preparación de las tesis, convirtiéndose así las publicaciones en un requisito previo para una evaluación positiva de cualquier programa de doctorado (Zamudio Igami, Bressiani y Mugnaini, 2014). Además, algunas universidades han comenzado a exigir a los candidatos al título de doctor la publicación de dos o tres artículos. En esta línea, los doctorandos de universidades españolas como Pompeu Fabra y Autónoma de Barcelona *han sido animados* a publicar los resultados de sus tesis en revistas revisadas por pares (Hollmann et al., 2015).

d) Los artículos derivados y los estudios bibliométricos

Se puede localizar término '*artículos de derivados*', referido a artículos basados en tesis, en estudios bibliométricos. Este término es empleado por Zamudio Igami, Bressiani y Mugnaini (2014) para denominar a aquellos artículos relacionados con tesis doctorales y publicados por los mismos autores, en un análisis sobre un programa de postgrado en la Institución de Investigación Nacional Nuclear de Brasil (IPEN-la CNEN / SP). Larivière también se refiere a "*documentos derivados de tesis*" (2010, p. 16) para describir un conjunto de publicaciones de los estudiantes de doctorado de la provincia de Quebec (Canadá) en su tesis doctoral: '*A bibliometric analysis of Quebec's PhD students*'

En resumen, la denominación '*obra derivada*' es una terminología específica empleada para referirse trabajos 'basados en' o 'extraídos de' obras previas. Explícitamente, '*artículo derivado de una tesis*' es empleado por editores de revistas, universidades e investigadores de bibliometría para referirse a aquellas publicaciones cuyos resultados de investigación proceden de tesis doctorales y son publicados en revistas revisadas por pares.

En general, la terminología '*artículo derivado*' no puede considerarse consistente, porque aunque su significado es explícito, la misma expresión puede tener diferentes matizaciones dependiendo de donde se enmarque el concepto.

## 2.2. Publicaciones sobre artículos derivados de tesis de medicina

<sup>1</sup> <http://www.apa.org/science/leadership/students/authorship-paper.pdf>

<sup>2</sup> <http://dkspublication.com/guidelines.html>

<sup>3</sup> <http://matjournals.com/Submission-Guidelines.html>

<sup>4</sup> <https://libraries.mit.edu/scholarly/publishing/publisher-policies-thesis-content-and-article-publishing/>

<sup>5</sup> <http://www2.le.ac.uk/library/downloads/copyright/Keeping%20Your%20Thesis%20Legal>

<sup>6</sup> <http://www.lib.umich.edu/copyright/publishing-your-thesis-proquestumi>

La literatura refleja un creciente interés por la productividad de las tesis doctorales. La publicación de los resultados de investigación de las tesis en revistas revisadas por pares, es considerado un indicador de su valor científico y facilita que la investigación sea conocida por la comunidad científica para su escrutinio y validación.

La identificación de la productividad, en términos de artículos publicados, ha sido estudiada a partir de dos objetivos generales, la actividad investigadora de los doctorandos y la evaluación de las tesis en función de los artículos derivados.

Un estudio del Observatoire des Sciences et Techniques (2002) estimó que la actividad investigadora de los estudiantes de doctorado representa entre el 10-20% de la investigación académica indexada en áreas científicas, técnicas y médicas (Scientific, Technical and Medical, STM). Un estudio más reciente realizado por Larivière (2011) determinó que las contribuciones de los doctorados representan alrededor de un tercio de la producción de las publicaciones en ciencias y medicina, aunque no se constataron relaciones de las publicaciones con las tesis doctorales.

Determinados estudios han investigado a partir búsquedas realizadas en bases de datos multidisciplinares (Web of Science, Scopus), bases de datos de medicina (Medline, Lilacs, SciELO) o buscadores académicos (Google Académico) las publicaciones generadas a partir de tesis de medicina. Los investigadores han intentado identificar las publicaciones mediante procedimientos de similitud textual basados en fragmentos cortos de texto, tales como palabras clave en el título y abstract, junto con el nombre del autor de la tesis, y en ocasiones, el nombre de los directores (Dhaliwal, Singh y Bhatia, 2010; Benotmane et al, 2012). Con el fin de identificar la producción derivada de la investigación doctoral (Zamudio Igami, Bressiani y Mugnaini, 2014) utilizaron un análisis de coocurrencia de términos, basado en pares de descriptores entre la tesis y los artículos, escritos por el mismo autor de la tesis.

Salmi, Gana y Mouillet (2001) realizaron un estudio sobre 300 tesis de medicina presentadas entre 1993 y 1998 en universidades francesas. Descubrieron que el 17% de las tesis dieron lugar a publicaciones indexadas. Benotmane et al. (2012) determinó a partir de 2.150 TEM (Thèses d'Exercice en Médecine) defendidas en la Facultad de Medicina de Lille 2, entre 2001 y 2007, que 243 TEM (11,3%) fueron objeto de alguna publicación científica, al tiempo que indicaron que "el número real de las publicaciones procedentes de tesis doctorales no es conocido, y por lo tanto, las comparaciones con otras facultades de medicina de Francia u otros países fue imposible " (2012, p. e398). La misma dificultad para determinar el número de publicaciones de programas de doctorado fue constatada en el Reino Unido por Caan y Cole (2012) al analizar publicaciones sobre investigaciones clínicas procedentes tesis doctorales de 39 universidades entre 2000-2010. Los resultados mostraron que el 47,6% de las tesis produjeron publicaciones no discernibles.

En Alemania, Cursiefen y Altunbas (1998) evaluaron la contribución de los médicos de la Universidad de Würzburg, entre 1993-1995. Los resultados revelaron que para un 66% de los estudiantes de medicina su investigación doctoral dio lugar a una publicación indexada en Medline. Por otra parte, Ziemann y Oestmann (2012), realizaron un estudio retrospectivo sobre artículos publicados por estudiantes de doctorado del Hospital Charité Universitätsmedizin de Berlín en la década 1998-2008. Los datos revelaron que el número de doctorados que habían publicado artículos había aumentado del 33% (1998) al 52% (2008). Sin embargo, más allá de estos datos cuantitativos Ziemann y Oestmann (2012) plantearon la cuestión de si *los proyectos científicos paralelos en el período de doctorado fueron documentados como publicaciones específicas relacionadas con el tema particular de las tesis*. En este sentido, un estudio sobre tesis españolas de anestesiología, llevada a cabo por Figueredo et al. (2002) reveló que 204 doctorandos produjeron un total de 1.679 artículos, de los cuales sólo 103 artículos (6,1%) estaba relacionado con la tesis. Dhaliwal, Singh y Bhatia (2010), basándose en criterios de autoría y en los resultados de investigación de las tesis, identificaron que la tasa de artículos

derivados de los autores de las tesis de una facultad de medicina de la India fue del 30%. Diferentes resultados fueron presentados por Frković et al. (2002) en relación con los artículos científicos basados en tesis doctorales correspondientes a dos facultades de medicina croatas, sugiriendo que las diferencias significativas entre las universidades (Rijeka 11,1%, Zagreb 40,9%) se debieron a la atención institucional orientada hacia publicaciones durante la etapa del doctorado.

Además, algunos estudios utilizaron encuestas como métodos complementarios para evaluar la validez de los resultados. De esta forma, Baufreton et al. (2012) llevaron a cabo un estudio de validación de los artículos derivados entre directores de tesis de la Escuela de Medicina de Angers (Francia), cuyos resultados mostraron que el 28% de las publicaciones procedían de tesis doctorales, de ellas 97 (16%) fueron indexadas en PubMed. Pitche, Onipoh y Tchangaï-Walla (2007) llevaron a cabo una encuesta entre los directores de tesis para conocer las publicaciones relacionadas con las tesis. El resultado fue que las tesis dieron lugar al 20,4% de las publicaciones de la Facultad de Medicina de la Universidad de Lomé (Togo). Zamudio Igami, Bressiani y Mugnaini (2014) utilizaron encuestas electrónicas enviadas a los autores de las tesis como método de validación, el resultado indicó que una media del 50% de los artículos publicados por los doctorandos estaba relacionada con la investigación doctoral.

Los estudios reseñados evidencian que sus resultados no son fácilmente comparables. Como indica Larivière la mayoría de los análisis se han realizado como *estudios de casos* a pequeña escala y "ningún estudio ha tratado de medir la producción de las tesis a nivel macro" (2011, p. 465). Esto no es sólo debido a los diferentes programas de postgrado, a los diferentes contextos multilingües donde los artículos no siempre son indexados en bases de datos, o a diferencias en las políticas de publicación de las instituciones, sino que la causa reside en la dificultad metodológica para determinar los trabajos publicados asociados a las tesis doctorales. Mientras que el aumento de las políticas de acceso abierto significa que el texto completo de las tesis se encuentra ahora más ampliamente disponible y el análisis se puede realizar a mayor escala, sin embargo hay cuestiones metodológicas que deben abordarse necesariamente para la identificación de estos artículos.

### **2.3. Perspectivas metodológicas para la identificación de artículos derivados**

Los estudios anteriormente reseñados han estudiado las publicaciones generadas a partir de las tesis de medicina utilizando diferentes enfoques metodológicos con resultados no fácilmente comparables. De hecho, la proporción de artículos relacionados con las tesis varía significativamente desde un 6,9% (Figueredo et al., 2002), 30% (Dhaliwal, Singh y Bhatia, 2010), 40,9% (Frković et al., 2002) a una media del 50% (Zamudio Igami, Bressiani y Mugnaini, 2014).

Las metodologías aplicadas por estos trabajos para detectar los artículos derivados se han basado fundamentalmente en la combinación del nombre del autor y/o nombre del director/es de las tesis con las palabras clave contenidas en los títulos y/o abstracts de los artículos.

- *Similitud textual*, palabras clave y/o descriptores contenidos en los títulos de los artículos y tesis, combinados con el nombre del autor de la tesis y/o director/es más el nombre los autores de los artículos (Figueredo et al, 2002; Frković et al, 2002; Zamudio, Igami, Bressiani y Mugnaini, 2014).
- *Similitud textual*, palabras clave contenidas en los títulos y abstract de los artículos y tesis, combinados con al menos, el nombre del autor de la tesis sea autor del artículo y que uno *de los resultados de la tesis* sea un resultado del artículo (Dhaliwal, Singh y Bhatia, 2010).

- *Encuestas* utilizadas para validar si los artículos son derivados, dirigidas a los autores de las tesis (Figueredo et al, 2002) o a los directores (Pitche, Onipoh, Tchangi-Walla, 2007; Baufreton et al, 2012), (Zamudio Igami, Bressiani y Mugnaini, 2014).

En contraste con esta perspectiva, otros investigadores mostraron la dificultad para distinguir las publicaciones derivadas de las tesis. Benotmane et al. señalaron que "la cifra de un 11,3% [artículos asociados a las tesis] podría parecer baja, debido a la falta de referencias para poder comparar datos con otras facultades de medicina de Francia o del extranjero" (Benotmane, 2012, E402). Caan y Cole (2012) señalaron que "39 de 87 tesis (47,6%) no produjeron publicaciones discernibles". Por otro lado, Ziemann y Oestmann (2012) cuestionaron si "los nuevos proyectos científicos paralelos considerados para este período [la presentación de tesis doctorales] se refieren a la actividad científica en general en la que se han documentado, en lugar de referirse a las publicaciones específicas relacionadas con el tema particular de las tesis" (Ziemann y Oestmann, 2012, p. 335).

En resumen, los diferentes intentos por obtener información sobre las publicaciones producidas por las tesis doctorales mostraron resultados no concluyentes, determinados por cuestiones metodológicas para la identificación de estas publicaciones.

### **2.3.1. Similitud textual basada en el contenido y estructura**

Con el objeto de obtener una mayor comprensión sobre los aspectos metodológicos en la identificación de los artículos derivados se examinaron desde una perspectiva de similitud textual enfoques sobre detección de plagio basados en el contenido y estructura de los artículos. En concreto, se examinaron estudios cuyas metodologías aplicadas para la detección del plagio se basaron en el texto completo de artículos de medicina y de esta forma poder descubrir posibles grados de coincidencia con esta investigación.

Bailey (2002) estudió los artículos publicados en las revistas de la *American Medical Association: Archives of Otolaryngology-Head y Neck Surgery*. La metodología consistió en agrupar los artículos, en primer lugar por similitud textual de los títulos, seguidos por la agrupación de las publicaciones de los autores en orden cronológico, para finalmente ser revisadas y analizadas en función de los siguientes conceptos: (i) similitud del contenido (ii) nivel de similitud textual. El análisis del texto completo reveló que el 20% de los autores había publicado artículos que contenían algún grado de duplicación y los tipos de duplicación más comunes incluían *datos* y *conclusiones* similares. De los 5 niveles de similitud considerados, el más frecuente correspondió a las publicaciones donde se había producido un aumento de la muestra analizada sin cambios en las conclusiones (41%). No hubo casos de duplicaciones definidas estrictamente en función de *párrafos idénticos reutilizados*. Hubo varios casos en los que *ilustraciones, leyendas y tablas* se repitieron. La reformulación y paráfrasis del texto fueron prácticas bastante comunes entre los autores que duplicaron trabajos.

Desde un punto de vista bibliométrico, García-Romero y Estrada-Lorenzo, (2014) analizaron artículos incluidos en la base de datos "Déjà vu", que contiene referencias de Medline altamente similares identificadas por el algoritmo eTBLAST. 'Déjà vu' establece una puntuación de similitud textual basado en los títulos y abstracts, combinados con el nombre de los autores y afiliaciones. Estos autores analizaron la similitud del texto completo de 247 pares de artículos de 'Déjà vu' que habían sido examinados por expertos siguiendo un estricto protocolo. Los resultados revelaron que la similitud encontrada en secciones de Introducción, Materiales y Métodos, podrían deberse a la publicación de estudios que formasen parte de la misma línea de investigación de un grupo científico. Sin embargo, consideraron que "los documentos que contienen una *gran similitud* en las secciones de los *Resultados* y *Conclusiones* deberían ser analizados antes de su aceptación por una revista".

El estudio sobre duplicados realizado por Errami et al. (2008) utilizó la base de datos 'Déjà vu'. La detección de similitud textual se basó en los abstracts y títulos de los artículos ya que "conceptualmente destilan los contenidos del texto completo". Entre las conclusiones apuntadas se encuentran, una alta proporción de *referencias duplicadas*, que en un 60%, son compartidas con el *artículo original* y, en consecuencia, "las referencias podrían ser usadas en el futuro para mejorar la sensibilidad en la detección de duplicados". Además, se hizo hincapié en las implicaciones que tienen las diferentes secciones de los artículos en la detección de duplicados, destacando que la *originalidad (novelty)*, que no es un componente fundamental de secciones como la Introducción o Materiales y Métodos, sí es parte integral de los *Resultados*, *Discusión* y *Conclusiones*, y por ello estas secciones "deberían ser considerada sacrosantas" (2008, p. 248).

Sun et al. (2010) utilizaron el algoritmo eTBLAST para localizar publicaciones biomédicas duplicadas en PubMed Central. (PMC). Para entender mejor los patrones de similitud textual entre las diferentes secciones de los artículos de PMC se computaron la Introducción, Métodos y Resultados/Discusión. Se halló que la similitud de los párrafos de la sección de Métodos fue aproximadamente 3,6 veces más frecuente que la de las secciones de la Introducción y 5,8 veces más frecuente que en las secciones de los Resultados/Discusión. Esto demostraría, utilizando el texto completo de los artículos, que la sección más proclive en ser reutilizada es la Metodología. Los autores calcularon también la frecuencia de similitud en un conjunto de artículos pares duplicados identificados por la base de datos de "Déjà vu". Los resultados obtenidos fueron que las secciones con el índice más alto similitud textual fueron los Resultados/[Discusión] (0,94), seguidos de la sección de Métodos (0,89) y la más baja la sección de Introducción (0,82). La comparación de los resultados obtenidos entre la PMC y el dataset de pares duplicados (Déjà vu') demostraría que, mientras que la similitud en la sección de Métodos generalmente es más común que la de otras secciones, la similitud que se produce en las secciones de Resultados / [Discusión] es *el mejor indicador de verdaderas publicaciones duplicadas*.

Por otro parte, desde un enfoque bibliométrico Larivière y Gingras (2010) analizaron artículos duplicados basados en los metadatos (título, autores y número de referencias). Los autores reconocieron que el método utilizado para identificar duplicados fue limitado debido a que los artículos duplicados modifican ligeramente sus metadatos no pudiendo ser capturados con esta metodología. Por lo tanto, los metadatos se consideraron insuficientes para detectar duplicados.

Los estudios mencionados parecen apuntar a una característica común en la identificación de plagio y la duplicación. El plagio inteligente de los artículos no se produce por la similitud textual de frases o párrafos, sino por la copia de '*las ideas creativas, aportaciones, resultados, discusiones y métodos de resolución de problemas*'. (Alzahrani, Salim y Abraham 2012a, p. 146). Como estos autores confirman el *plagio está determinado por la copia de las ideas* a través de las secciones que son segmentos substanciales del trabajo científico como son, *los Resultados, Discusiones, Hallazgos y Contribuciones*, no por unas pocas palabras prestadas.

Desde un punto de vista retórico, el contenido informativo de las secciones [Resultados/ Discusión] forman el núcleo informativo de los artículos porque:

- Resultados, presentan el nuevo conocimiento y por lo tanto el valor intrínseco del artículo depende de su contenido.
- Discusión, es la sección donde los autores explican el significado y las implicaciones de los Resultados, con el apoyo y contraste de la bibliografía. La redacción de esta sección enfatiza los nuevos hallazgos contrastándolos con investigaciones previas, en términos de importancia y valor.

### **2.3.2. Similitud textual basada en las citas bibliográficas**

La reutilización de las publicaciones científicas disponibles en Internet sin citar a los autores originales es una práctica común considerada plagio. Varios estudios se han referido a la prevalencia del plagio en las publicaciones médicas (Schein y Paladugu, 2001; Von Elm et al., 2004; Errami et al., 2008; García-Romero y Estrada-Lorenzo, 2014). Dada la importancia de los casos de plagio en la comunidad científica, numerosos investigadores se han dedicado a diseñar sistemas para automatizar su detección en base al contenido y a las referencias bibliográficas.

Esta parte revisa los experimentos realizados utilizando las referencias bibliográficas como fuente de similitud. El enfoque de usar las referencias es reciente y hay poca evidencia empírica sobre su eficacia. Los experimentos llevados a cabo analizaron bien las referencias comunes como filtro para detectar los falsos positivos (casos de no plagio) Sorokina et al., (2006), o una colección de referencias como método de detección de documentos sospechosos. Alzahrani, Salim y Abraham (2012a).

El uso de las referencias como fuente de similitud para detectar el plagio en las tesis doctorales fue analizado por Gipp (2013) en su propia tesis doctoral '*Citation-based plagiarism detection: applying citation pattern analysis to identify currently non-machine-detectable disguised plagiarism in scientific publications*' (Universität Magdeburg). Gipp (2013) indicó que a diferencia de los enfoques que realizan comparaciones entre contenidos textuales, los patrones de las referencias bibliográficas se pueden utilizar como "*huella digital semántica*" (semantic fingerprint) para identificar potencialmente la similitud entre documentos. La idea que subyace en esta investigación es que quienes plagian, comúnmente disfrazan el texto copiado parafraseándolo, pero no sustituyen o reorganizan las referencias de forma significativa.

Los algoritmos empleados por Gipp y Beel (2010), Gipp, Meuscheke y Beel (2011) Gipp, (2014) para la detección de modelos basados en las referencias bibliográficas dependen de marcadores cuyas características son la proximidad, superposición, orden, frecuencia y carácter distintivo. El procedimiento para la detección del plagio contiene diferentes fases, (i) la aplicación de los algoritmos a las referencias bibliográficas para obtener un conjunto de datos, (ii) detección de falsos positivos y (iii) llevar a cabo un estudio de validación por expertos respecto a la relevancia de los documentos sospechosos de plagio.

Pertile, Rosso y Moreira (2013) realizaron un estudio consistente en comparar la similitud de trabajos científicos basándose en el análisis de coocurrencias de las referencias bibliográficas. La hipótesis fue que una alta tasa de coocurrencias de referencias entre documentos podría ser un indicador de plagio. Los resultados indicaron que en la mayoría de los casos donde se encontraron coocurrencias en las referencias se correspondieron de hecho con pasajes textuales plagiados.

### **2.3.3. Comparación basada en contenido y citas**

Diferentes enfoques han investigado hasta qué punto los métodos de detección basados en el contenido y referencias bibliográficas pueden complementarse y mejorar los enfoques individuales. Ahlgren y Colliander (2009) midieron la similitud entre documentos comparando el texto de los contenidos y referencias. Los resultados revelaron que los métodos basados sólo en las referencias obtuvieron peores resultados que los métodos basados sólo en el texto de los contenidos.

Boyack y Klavans (2010) compararon tres técnicas basadas en referencias puras, *direct citations*, *bibliographic coupling*, *co-citation analysis* y *citation text hybrid approach*<sup>7</sup> con objeto de seleccionar el método que podría adaptarse mejor la investigación en biomedicina. Los resultados indicaron que el enfoque híbrido de referencias y texto (*citation text hybrid approach*) supera a otros enfoques. Varios estudios (Glenisson, Glänzel y Persson, 2005; Janssens et al, 2006; Ahlgren y Colliander, 2009) han confirmado que la combinación del análisis de similitud del texto completo y las referencias mejoran a los métodos individuales.

Pertile, Moreira y Rosso (2015) llevaron a cabo un primer análisis basado en la correlación entre métricas de similitud textual del contenido y de referencias bibliográficas. Para evaluar la exactitud de los datos se calcularon los valores de *precision*, *recall*<sup>8</sup>, *F1*<sup>9</sup> y *mean average precision (MAP)*<sup>10</sup>. Los casos identificados como verdaderos positivos (VP) fueron contrastados con la prueba *Ground truth*<sup>11</sup>. Los autores indicaron que "los datos mostraron que los análisis basados en el contenido tienen un mejor rendimiento que los basados en otras métricas. Entre las diferentes métricas utilizadas para medir la similitud de las referencias la más precisa fue *bibliographic coupling* (2015, p. 12). Los autores realizaron un segundo análisis combinando las métricas basadas en el contenido y las citas. Los resultados obtenidos mostraron ligeras ventajas comparadas con el uso de las métricas individualmente.

Echeverría, Stuart y Blanke 2015b) (Véase Publicaciones) compararon la relación entre la similitud textual basada en el contenido y las referencias bibliográficas sobre una muestra de 68 tesis doctorales de medicina y 324 artículos publicados por los mismos autores de las tesis doctorales. La métrica utilizada para medir la similitud de las referencias fue *bibliographic coupling*. Los resultados obtenidos sobre la predictibilidad de las referencias comparada con la similitud textual de las secciones discursivas de los artículos (Introducción, Metodología, Resultados y Discusión) indicaron que la sección de las referencias tiene una alta sensibilidad para detectar casos verdaderos positivos y una baja especificidad para identificar casos verdaderos negativos, mientras los resultados basados en la similitud textual de la sección de Discusión revelaron una menor sensibilidad para detectar verdaderos casos positivos, pero una mayor especificidad para detectar verdaderos casos negativos [Véase Resultados, 4.3.3]

Estos análisis mostraron que el enfoque de similitud textual basado en el contenido supera en precisión al método basado en las referencias para detectar similitud entre documentos.

## 2.4. Indicadores bibliométricos de publicaciones

La bibliometría es el análisis cuantitativo de las publicaciones que permite medir de manera objetiva los documentos científicos desde tres perspectivas: la producción científica, que contabilizada el número de publicaciones de un autor o grupo de investigación, el impacto de las publicaciones en la comunidad científica y la visibilidad de las publicaciones.

La aparición de la bibliometría como campo de investigación fue iniciada en el Institute for Scientific Information (ISI) en 1960 por Eugene Garfield. El indicador desarrollado por ISI más conocido es el "Factor de Impacto de Revistas" (JIF), diseñado para medir las publicaciones

---

<sup>7</sup> *Direct citation*, cuando los artículos están vinculados porque uno hace referencia a otro, *bibliographic coupling*, cuando los documentos enlazados comparten referencias comunes y se define como 'la frecuencia con la que dos documentos citan otros documentos comunes' y *citation text hybrid approach*, enfoque que combina tanto métodos de similitud textual como referencias.

<sup>8</sup> *Recall*, recuperación exhaustiva

<sup>9</sup> *F1*: medida definida por Van Rijsbergen (1997) que combina *recall* (R) y *precision* (P).

<sup>10</sup> *Mean Average Precision (MAP)*, conjunto de puntuaciones medias de *precision*.

<sup>11</sup> *Ground truth*, método basado en la observación directa en contraste con la información obtenida por inferencia.

científicas de acuerdo con su impacto en la comunidad científica. La idea subyacente de este indicador, relevante todavía hoy, como señala Thelwall es su capacidad “para conocer qué artículos han citado un trabajo, ya que probablemente traten temas similares y puedan actualizar o corregir al artículo original ”(2008, p. 2).

La aplicación de herramientas bibliométricas requiere algunas consideraciones como explica Sanz Casado (1997). Los indicadores bibliométricos describen aspectos concretos de un estudio, y sólo el uso de varios indicadores permite obtener una buena visión general tanto los autores como de los recursos de información estudiados. Asimismo, los indicadores tienen un carácter relativo ya que “se refieren sólo al tipo de centro de información donde están siendo aplicados o al colectivo de usuarios estudiados” (1997, p.46).

Esta sección revisará tres tipos de indicadores:

1. Factor de Impacto
2. Citas recibidas
3. Indicadores de colaboración

### **2.4.1. Factor de Impacto**

El Factor de Impacto (FI) se define como el número de citas indexadas por el ISI para los artículos publicados por una revista en el año X, contabilizando las citas recibidas por los artículos publicados en los dos años anteriores, dividido por el número total de documentos académicos citables publicados en esos mismos años.

A pesar de su amplio uso como una herramienta de análisis y evaluación se le han atribuido algunas limitaciones :

- *Asimetría.* Archambault y Larivière (2009) observaron que hay asimetría entre las citas contabilizadas en el numerador e ítems contabilizados en el denominador. En efecto, mientras que las citas recibidas por todos los tipos de documentos publicados por la revista (artículos, notas y comentarios, así como editoriales, cartas al director, noticias, etc.) se contabilizan en el numerador, sólo los artículos, notas y artículos de revisión son considerados como documentos citables en el denominador. Por lo tanto, una revista que publica muchos editoriales y cartas al editor aumentará sus citas y factor de impacto en comparación con las revistas que sólo publican artículos.
- *Autocitas de revistas.* Garfield y Sher (1963) decidieron incluir en el análisis bibliométrico las autocitas de las revistas. Sin embargo, no previeron que su elección, abrió el camino a los editores para la manipulación del FI ‘*animando*’ a los autores a citar artículos de la revista *en la que trataban de ser publicados*. A pesar de ello, los estudios sobre el FI demuestran que la tasa de las autocitas de revistas es baja en cuanto al incremento del FI (Campanario y Molina 2009). La base de datos Web of Science (WoS) con el fin de mantener la transparencia del FI publica anualmente *Titles suppression from Journal Citation Report*, uno de cuyos criterios para la exclusión de una revista es el abuso de autocitas de revistas individuales. Además, las nuevas métricas como Eigenfactor y Article Influence no tienen en cuenta las citas procedentes de las autocitas de revistas.
- *Ventana de citación de dos años.* El período de dos años para la contabilización de las citas, fue adoptado por Garfield y Sher (1963) a principios de la década de 1960. Su creación obedeció a razones prácticas de tiempo y coste para la adquisición de datos que permitieran seleccionar las revistas de la publicación *Current Contents* (Larivière, 2009). No obstante, hay un consenso generalizado que considera el plazo de citación de dos años demasiado corto, especialmente para las disciplinas de Ciencias Sociales y Humanidades.

Posteriormente a su introducción, Garfield escribió un par de artículos (Garfield, 1998; Garfield, 2000) donde trató de justificar la ventana de citación de dos años en función del número importante de documentos que se citan de forma rápida y cuya confirmación completa se obtiene generalmente dos años más tarde de su publicación. Actualmente, WoS ha añadido una nueva métrica '*Five-Year Impact Factor*', cuya ventana de citación es de 5 años.

- *Distribución de las citas.* Las citas recibidas por los artículos publicados en una revista siguen una distribución no-paramétrica como indicó Seglen "más del 15% de los artículos contabiliza el 50% de las citas, y el 50% de los artículos más citados representan el 90% de las citas" (1997, p. 499). En otras palabras, un pequeño número de artículos altamente citados puede influir en el FI de una revista. Por tanto, el FI de una revista sería un mal predictor de las citas recibidas por los artículos.

La aplicación del FI en decisiones de evaluación, promoción y rendimiento de la investigación ha demostrado ser poco consistente como han señalado algunos investigadores (Garfield, 1998; Bordons, 2002; Torres-Salinas, 2007; Thelwall, 2008) ya que el FI se emplea como medida de las *citas esperadas* para cada uno de los artículos publicados, lo que indirectamente representa una medida sobre su posible calidad e impacto. Por el contrario, como demostró Seglen (1997) la correlación entre las *citas esperadas* y *observadas* es asimétrica.

A pesar de estas limitaciones mencionadas del FI, esta métrica sigue siendo utilizada ampliamente entre los investigadores de bibliometría debido a su gran accesibilidad, y razones de comparabilidad con la mayoría de los estudios bibliométricos existentes.

#### **2.4.2. Indicadores basados en citas**

Las citas recibidas son una medida fundamental utilizada en bibliometría para evaluar el impacto de las publicaciones. Tradicionalmente, el número de citas recibidas por un artículo ha sido visto como un marcador de la importancia e impacto de la investigación. Investigadores de bibliometría se han interesado en conocer los hábitos de citación y las principales razones que inducen a los autores a citar documentos. En esta línea, es un bien conocido el esquema prescriptivo diseñado por Garfield sobre cuándo citar "*When to cite*" (Garfield, 1996). La recomendación general se basa en que los autores deben citar a otros como "reconocimiento de deudas intelectuales". Sobre esta base, las citas deberían ser utilizadas para:

- Reconocer trabajos relacionados
- Identificar trabajos que '*deben*' ser citados
- Reconocer la calidad de una determinada obra

Según estas recomendaciones, las citas son vistas como un reconocimiento al que están obligados los autores. Esto significaría que el número de citas que uno recibe sería directamente proporcional al reconocimiento recibido. Esta perspectiva de Garfield sobre la citación está influenciada por las teorías de Merton (1973) sobre la sociología de la ciencia. Merton abordó la ciencia como una institución social estructurada sobre normas que caracterizan el comportamiento de los científicos. Expone que la publicación y citación certifican el descubrimiento científico y su reconocimiento por la comunidad científica. De esta forma, el reconocimiento es un elemento de la institucionalización de la ciencia que tiene como finalidad asignar una especial importancia al mérito y originalidad científica en el plano social.

Numerosos investigadores han negado este argumento normativo del reconocimiento científico de las citas como una medida de la calidad basada en la evidencia empírica. Seglen considera que "el criterio principal para la selección de una cita es la utilidad dentro de la investigación,

en lugar de la calidad científica pura" (1997, p. 1050). Higgins y Caso (2000) proporcionan una sinopsis sobre la motivación de los autores en el uso de citas, tales como la probabilidad de citar artículos altamente citados, citas de revisión, trabajos que discuten aspectos metodológicos, obras consideradas estudios clásicos y *citas como resultado de la relación entre el autor que cita y autor citado*. Borgman (2001) describe que la motivación y objetivos de los autores que citan son complejos, y en ellos pueden influir más los intereses profesionales que los argumentos puramente normativos de Garfield. Como síntesis argumenta que no todas las obras que *'deben'* ser citadas lo son, y no todas las obras que se citan *'deben ser citadas'*.

Hace tiempo que se reconoce que "no todas las citas son iguales" Wouters, (1999) y, por tanto, no deberían contarse por igual. La pregunta clave es cuál es la correlación entre la calidad y la cantidad. En este sentido, se han planteado algunas métricas como "Article Influence" y "Eigenfactor" (Bergstrom y West, 2008) [Véase 3.8.1.2, 6.1.2.3 y 6.1.2.4]. Su cálculo se basa en el número de veces que los artículos han sido citados y en la importancia de las revistas que han contribuido a estas citas. El enfoque basado en ponderación de las citas propone que las revistas más altamente citadas reciben más ponderación.

En conjunto, los estudios demuestran que el análisis de las citas como procedimiento para evaluar el impacto de una investigación puede verse afectado por diferentes factores que incluyen desde la calidad y el enfoque del contenido (Johnston, Piatti y Torgler 2013), la disponibilidad de la publicación, la influencia de los autores (Merton 1968), el factor de impacto de la revista, el número de autores por artículo, y un componente más difícilmente detectable como son las citas sociales (Cole y Cole, 1967).

A pesar de la importancia que tienen las citas para determinar el impacto de una obra en la literatura científica, la evaluación de la calidad de una investigación en un campo científico es un aspecto que sólo podría ser juzgado por expertos en la materia.

### **2.4.3. Indicadores de colaboración**

#### **2.4.3.1. Colaboración de los directores**

El director de una tesis es el enlace principal entre el doctorando, el departamento y la institución. La integración del doctorando en un equipo de investigación se establece a través del director, quien pone en relación al estudiante con el equipo en múltiples aspectos: teóricos, metodológicos y de capital social.

La figura del director está justificada no sólo por la necesidad del estudiante de aprender habilidades básicas y avanzar en su investigación, sino también porque le proporciona una ruta a través de la cual el estudiante puede integrarse en la institución, pudiendo llegar a ser este esencial en el progreso de su doctorado. La literatura indica una amplia gama de actividades y responsabilidades realizadas por los directores (Pole et al., 1997), desde funciones puramente académicas o técnicas, a funciones orientadas al desarrollo de la investigación y de apoyo en proyectos científicos, aspectos considerados básicos para el progreso del doctorado.

La influencia del director en la productividad de los doctorandos ha sido estudiada desde diferentes perspectivas. La presencia de los directores en las publicaciones relacionadas con las tesis fue analizada por Zamudio Igami, Bressiani y Mugnaini (2014). Los resultados mostraron una relación directa entre la presencia de los directores como coautores y los *'artículos relacionados'* con la tesis (70,60%), mientras que en el caso de los artículos *'no relacionados'* con las tesis la participación de los directores como coautores fue menos relevante (47,9%).

Pole *et al.* (1997) y Larivière (2011) estudiaron la presencia de directores/mentores como coautores en las publicaciones de los doctorandos. Sugirieron que, incluso si no se dispusiera de información sobre la relación entre el doctorado y otros autores, se podría esperar, que los coautores serían sus directores/mentores.

Kamler (2008) exploró desde una perspectiva educativa el papel de los directores en las publicaciones de los doctorandos. El análisis confirmó un aumento de publicaciones en aquellos casos donde los doctorandos habían recibido la supervisión de los directores en términos de revisiones críticas, mejora en el proceso de escritura o estrategias orientadas a la publicación. Por ello, el autor reconoce la importancia de articular una atención educativa a los doctorandos dentro de una política coherente en materia de publicación de „, dentro de la cual los directores jugarían un papel significativo de ayuda a los doctorandos tanto en aspectos científicos como educativos.

Golde definió la influencia del apoyo intelectual e interacción con los directores como el componente central de la educación doctoral y un factor clave donde están unidos “el proyecto de investigación, la financiación, el patrocinio profesional y las publicaciones” (2005, p. 287).

Los efectos del apoyo del director en las publicaciones de las investigaciones doctorales fueron analizados por Dinham y Scott (2001) a través de una encuesta con participantes de EE.UU. y Canadá. Sus conclusiones fueron que los estudiantes de doctorado que recibieron algún tipo de estímulo o ayuda por parte de los directores fue del 66% frente al 34% que no recibieron esta ayuda. Los efectos de esta ayuda se tradujeron en que aquellos doctorandos que habían recibido el apoyo de los directores habían publicado un 77% antes de la graduación y un 91% después de la graduación. Por el contrario, los doctorandos que no recibieron dicho apoyo publicaron antes de la graduación un 44% y un 50% después su graduación.

En general, la literatura refleja el papel clave desempeñado por los directores en relación con la producción publicaciones, en la finalización las tesis con éxito y en la participación de los estudiantes en las redes académicas. Sin embargo, el grado de participación y supervisión que llevan a cabo los directores no es fácil conocer, debido al déficit de información objetiva sobre los estudiantes de doctorado y a la falta de información sobre las actividades internas de los departamentos.

#### **2.4.3.2. Colaboración de los doctorandos en equipos de investigación**

La investigación científica se está convirtiendo en una tarea cada vez más colaborativa (Subramanyam, 1983). La colaboración científica es una característica inherente de la investigación médica y tiene efectos positivos sobre la productividad, la visibilidad y el impacto de la investigación (Zuckerman, 1968; Bordons et al, 1996; Glänzel, 2001).

El factor más objetivo de la integración estudiantes de doctorado en la comunidad académica es su participación como autores de publicaciones científicas. Subramayan indica que "la frecuencia relativa de coautoría en grupos [de investigación] es proporcional al grado de colaboración científica" (1983, Pág. 35).

El contexto de integración de los doctorados en el área biomédica se produce en los laboratorios, departamentos o a nivel de equipos de investigación. En este entorno, los estudiantes de doctorado pueden llegar a tener un papel relevante, lo que significa que participan en los proyectos de investigación cuya gestión y propiedad pertenece al equipo. Los doctorandos de medicina además de estar implicados en la propia investigación de su tesis, suelen colaborar en otras investigaciones del departamento o del laboratorio con otros estudiantes, becarios postdoctorales, o profesores. De este modo, los doctorandos se convierten en una parte

sustancial de los equipos de investigación y sus propias investigaciones, con frecuencia, se construyen dentro de las mismas líneas de investigación del departamento.

Larivière (2011) analizó el resultado de las investigaciones realizadas por los estudiantes de doctorado de Medicina, Ciencias e Ingeniería en Quebec (Canadá). Su estudio reveló que la proporción de doctorados como autores de al menos un artículo fue del 63% en Medicina y un 40% en Ciencias e Ingeniería. Los datos revelaron que los doctorandos llegan a ser *de facto* una parte integral del trabajo de los equipos de investigación.

A pesar del papel desempeñado por los doctorandos en las publicaciones científicas todavía no se ha abordado un análisis desagregado sobre su colaboración con otros autores: a nivel de la misma organización (intra-institucional), de otras instituciones del mismo país (colaboración inter-institucional) o con autores que trabajan en el extranjero (colaboración internacional), debido a la dificultad de discernir el estatus individual de los autores y la posición que ocupan.

En conclusión, la participación de los estudiantes de doctorado en la investigación es un componente importante de las publicaciones académicas porque como demuestra Larivière "la contribución de los estudiantes de doctorado representa aproximadamente un tercio de la producción de publicaciones" (2012, p. 463), lo que evidencia el alto grado de su participación en equipos de investigación.

## **2.5. Factores extrínsecos relacionados con las publicaciones**

### **2.5.1. Políticas de doctorado**

La comunicación del conocimiento es una parte esencial del proceso de investigación. Los resultados científicos no son parte del discurso científico hasta que son publicados y sometidos al análisis crítico de los revisores. Tradicionalmente, las revistas han desempeñado este papel fundamental de difusión y certificación de los nuevos conocimientos.

En base a estas premisas, las políticas recientes de las universidades están intensificando el requerimiento de solicitar publicaciones a los estudiantes de doctorado, se les anima a publicar durante, y como parte de doctorado (Kamler, 2008). Las políticas que impulsan la publicación o las políticas de mandato de publicación parecen existir en muchas instituciones, hecho que contradice las observaciones de Dinham y Scott cuando afirman que "las políticas que obligan a la publicación parecen existir en una minoría de las instituciones" (2001, p. 53).

A nivel internacional, ha habido una serie de iniciativas para promover la difusión de los resultados de las tesis en las publicaciones. Políticas como las promovidas por la *European University Association* han favorecido la importancia de las publicaciones durante la etapa del doctorado "Los programas de doctorado (...) deben buscar la independencia de la investigación y la publicación de resultados" (2005b, p. 25).

En el Reino Unido, las presiones por publicar antes de la defensa de la tesis están desafiando el modelo de tesis tradicional basada en una monografía por capítulos. Una encuesta realizada entre 50 universidades del Reino Unido, llevada a cabo por Christianson, Elliot y Massey (2015) reveló que en las disciplinas del área científica normalmente se espera que un estudiante de doctorado haya publicado algún artículo antes de la lectura de la tesis. Varias instituciones conscientes de que las publicaciones que integran las tesis podrían incrementar el tiempo de su preparación han establecido como incentivos que los candidatos reutilicen el material publicado incorporándolo como parte integral a la tesis, sin reescribirlo. Además, un número importante

de instituciones están considerando un cambio en el modelo de las tesis donde las publicaciones se conviertan en una parte substancial y otras están considerando la posibilidad de abordar el formato de las *'tesis por publicación'* para todos los candidatos, en lugar de limitarlo a miembros del personal académico (Christianson, Elliot y Massey, 2015).

En España, según el Real Decreto 1393/2007 sobre educación universitaria se establece que "uno de los principales fundamentos del sistema de educación superior es la investigación, así como la difusión de los resultados generados". Si bien no es obligatorio para los estudiantes de un máster realicen publicaciones, los programas de doctorado promueven la publicación de los resultados de investigación, teniendo en cuenta los beneficios académicos y profesionales derivados.

Se examinaron algunos programas de doctorado de universidades españolas respecto a la política de publicaciones durante el periodo doctoral con el siguiente resultado:

- Universidad Pompeu Fabra y Universidad Autónoma de Barcelona. Una encuesta realizada por Hollmann et al. (2015) mostró que el 62,5% de los estudiantes habían publicado o estaban en proceso de publicación de *'un artículo derivado de su tesis'*.

-Universidad Politécnica de Madrid<sup>12</sup>. El programa de doctorado señala que "para que una tesis doctoral pueda ser defendida, deberá contar con un avance de los resultados difundidos en una publicación con evaluación previa".

- Universidad Politécnica de Valencia<sup>13</sup>. El programa de doctorado de esta universidad indica que "La presentación de una tesis doctoral será aceptada cuya memoria incluya directamente artículos en revistas indexadas u otras publicaciones de reconocido prestigio sometidas a revisión por pares, cuyo autor destacado sea el doctorando".

-Universidad Nacional de Educación a Distancia (UNED)<sup>14</sup>, de acuerdo con los requisitos para la presentación de tesis doctorales "los estudiantes necesitarán justificar, documentalmente, haber realizado alguna publicación o patente relacionada con el tema de investigación de su tesis, o contar con la aceptación de los editores para la publicación del trabajo".

Las políticas mencionadas en los programas de doctorado demuestran que las universidades españolas están impulsando abiertamente la publicación de *artículos derivados de tesis* en revistas revisadas por pares.

## 2.5.2. Políticas sobre tesis electrónicas

Existe un acuerdo implícito entre el personal académico sobre las ventajas derivadas de disponibilidad de las tesis en OA, ya que potencia su visibilidad a través de los repositorios, Google Scholar o herramientas descubrimiento, y facilita que las tesis puedan ser más ampliamente utilizadas, visitadas, leídas y citadas (Troman, Jacobs y Copeland, 2007; Cullen, Chawner, 2011).

---

<sup>12</sup> Universidad Politécnica de Madrid. *Elaboración y evaluación de la tesis doctoral*. <http://www.upm.es/sfs/Rectorado/Vicerrectorado%20de%20Doctorado%20y%20Postgrado/Negociados%20de%20Doctorado%20y%20Postgrado/Impresos/Reglamento%20de%20Elaboraci%C3%B3n%20y%20Defensa%20de%20Tesis%20UPM.pdf>.

<sup>13</sup> Universitat Politécnica de Valencia. *Descripción detallada del procedimiento de elaboración y defensa de tesis doctorales*. <http://www.upv.es/entidades/SA/tercerciclo/defensa.pdf>

<sup>14</sup> Universidad Nacional de Educación a Distancia (UNED). *Requisitos para la presentación de tesis doctorales. Acreditación de publicaciones y patentes*. [http://portal.uned.es/portal/page?\\_pageid=93,25486040&\\_dad=portal&\\_schema=PORTAL](http://portal.uned.es/portal/page?_pageid=93,25486040&_dad=portal&_schema=PORTAL)

Sin embargo, y particularmente los estudiantes de doctorado y directores de tesis han expresado una gran preocupación por las implicaciones de las ETDs (Tesis Electrónicas y Disertaciones) en OA. Entre las preocupaciones más comúnmente citadas se encuentra las potenciales limitaciones sobre la capacidad de publicar los contenidos de las tesis en un libro o en una serie de artículos de revistas, así como la posibilidad de solicitar el reconocimiento de patentes cuando el pre-print (tesis) está accesible en OA un repositorio (Greig, 2005; Stanton y Liew, 2011).

Otro tema relacionado con la publicación en OA es el copyright. En efecto, existe el riesgo de un mal uso de la tesis en OA, no sólo para su autor, sino también para los titulares de derechos de cualquier material incluido en el contenido de la tesis (Barnes et al., 2012). Otra preocupación común surge respecto al riesgo de plagio, ya que si los datos y los resultados de una investigación se publican en OA antes de su publicación formal, pueden dar lugar a una ventaja competitiva de los hallazgos científicos a potenciales rivales (Stanton y Liew, 2011). Además, el personal académico ha expresado confusión y preocupación por los derechos de autor cuando los materiales académicos son reutilizados creándose nuevas obras a partir de otras anteriores (Gadd, Loddington y Oppenheim, 2007).

Un problema particularmente asociado a las tesis en OA es el '*efecto potencial*' sobre las futuras oportunidades de publicación. Virginia Tech fue la primera universidad en exigir que las tesis se presentaran en formato digital en 1997. Desde el principio, profesores y estudiantes expresaron que la publicación de las ETDs en OA podría ser interpretado como *prior publication*, y en consecuencia se podría comprometer la publicación del material de la tesis en una revista, un libro o patentes pendientes de reconocimiento, así como se podría revelar planes futuros de investigación (Thompson, 2001). Estas cuestiones provocaron actitudes en contra de la publicación de las tesis en OA, obligando a la universidad a crear diferentes niveles de acceso a las ETDs.

Respecto a las actitudes de los estudiantes y directores sobre las ETDs en OA, Brown (2010) llevó a cabo una revisión sobre este tema en diferentes países. Destaca que, aunque existe una clara mayoría de estudiantes de postgrado que perciben los beneficios las ETDs, diseminación de contenidos, incremento de acceso, rápida distribución y posibilidad de recibir más citas, paralelamente también manifiestan una serie de preocupaciones, predominantemente sobre el impacto sobre futuras publicaciones, derechos de autor, plagio y el material sensible del contenido de las tesis, proponiendo que las ETDs puedan ser gestionadas mediante un procedimiento de depósito correctamente diseñado.

Además, es significativo que defensores del OA como Suber sugieran explícitamente que "[las universidades] exijan a los estudiantes que publiquen los artículos antes de terminar su tesis. Los estudiantes que quieran utilizar los artículos en sus tesis deberían conservar el derecho de archivo en OA. En el caso de estudiantes que fallesen en mantener los derechos de archivo de los artículos podrían ser obligados a retrasar su publicación en una revista hasta que no hubieran aprobado la tesis doctoral" (2008, p. 32).

En la misma línea, Harnad defiende que las tesis y disertaciones son una categoría de investigación académica y científica que podría constituir una excepción al mandato de autoarchivo en OA, específicamente cuando los autores de las tesis esperan publicarlas como libros y obtener un beneficio económico. De ahí que puedan temer que la publicación de la tesis en OA haga imposible encontrar un editor, o reducir sus perspectivas de ingresos. "Estos autores pueden o no tener razón respecto a sus temores, pero es evidente que constituyen una excepción a la regla de OA". (2009, p. 2).

### 2.5.3. Políticas de editores de revistas y tesis electrónicas

Con el fin de inhabilitar la publicación de obras que habían circulado como preprints *New England Journal of Medicine* adoptó en 1969 la política de publicación conocida como la ‘Norma Ingelfinger’, que establece que las revistas sólo considerarán para su publicación los manuscritos cuya contenido substancial no se hubiera presentado o publicado en otra parte.

La aplicación de la ‘Norma Ingelfinger’ ha sido objeto de discrepancias entre los editores de revistas respecto a la publicación de trabajos basados en tesis en OA. El componente clave considerado por los editores es si las ETDs en OA constituyen *prior publications*. En este sentido, McMillan et al. (2011) realizaron una encuesta entre editores de revistas en 2011 y sus resultados fueron comparados con dos encuestas anteriores realizados por Dalton y Seamans (2001) en el año 2000 y Seamans (2003) en el año 2001. La comparación de los resultados de las encuestas fue la siguiente:

	2000 Dalton	2001 Seamans	2011 NDLTD
Welcomed for submission	47%	33%	45%
Case by case basis	19%	29%	27%
Contents substantially revised	6%	6%	13%
ETD had limited access	6%	0%	3%
Under no circumstances	4%	2%	4%
Other	19%	27%	9%

Tabla 1. ‘How do 2011 findings compare to previous surveys.  
En: *An investigation of ETDs as prior publications (2011)*

La tabla refleja que las actitudes de los editores no han cambiado mucho respecto a las ETDs en 10 años. En general, la mitad de los editores aceptarían para su publicación los trabajos derivados de las ETDs. Además, un número relativo de editores, en torno al 20%, consideraría que la publicación de los resultados de las ETDs *caso por caso* y una minoría, en torno al 3%, no aceptaría la publicación los resultados de las ETDs.

Estas encuestas revelaron una cuestión básica, sobre la diferencia entre una tesis y obra derivada de una tesis, independientemente de si el trabajo está en formato electrónico o en papel. Los resultados de la encuesta indicaron que la mayoría de los editores rara vez publicarían una tesis que no hubiera sufrido alguna revisión, y por tanto, daría lugar a una *obra derivada* (Seamans, 2003).

En general, la mayoría de los editores manifestaron que no disponían una política definida respecto a las publicaciones derivadas de las tesis doctorales en OA. Por otra parte, la influencia de la regla Ingelfinger sigue siendo importante en los principales títulos de las revistas biomédicas (*New England Journal of Medicine*, *JAMA*, *Lancet*) y por extensión a las revistas de medicina. Según esta Regla, y con excepción de las presentaciones limitadas en reuniones científicas, las obras que han sido previamente publicados no serían consideradas para su publicación, disuadiendo de esta forma severamente la circulación y publicación de documentos en Internet (Armstrong, 2014).

## **8. BIBLIOGRAFIA**

## BIBLIOGRAFIA

- Aguillo, I.F. (2011), "Is Google Scholar useful for bibliometrics? A webometric analysis", *Scientometrics*, Vol. 91, No. 2, pp. 343-351.
- Ahlgren, P. y Colliander, C. (2009), "Document–document similarity approaches and science mapping: Experimental comparison of five approaches", *Journal of Informetrics*, Vol. 3, No. 1, pp. 49-63.
- Aksnes, D.W. (2003), "Characteristics of highly cited papers", *Research Evaluation*, Vol. 12, No. 3, pp. 159-170.
- Alzahrani, S.M., Salim, N. y Abraham, A. (2012a), "Understanding plagiarism linguistic patterns, textual features, and detection methods", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, Vol. 42, No. 2, pp. 133-149.
- Alzahrani, S., Palade, V., Salim, N. y Abraham, A. (2012b). "Using structural information and citation evidence to detect significant plagiarism cases in scientific publications". *Journal of the American Society for Information Science and Technology*, Vol. 63, No.2, 286-312.
- American Psychological Association (2015), "Ethics code updates to publication manual". Disponible en: <http://www.apa.org/ethics/code/manual-updates.aspx> [Consultado 2015, 27 de Julio].
- Andrew, T. (2004), "Intellectual property and electronic theses, JISC Legal". Disponible en: [https://www.era.lib.ed.ac.uk/bitstream/handle/1842/612/IP\\_etheses.pdf?sequence=2&isAllowed=y](https://www.era.lib.ed.ac.uk/bitstream/handle/1842/612/IP_etheses.pdf?sequence=2&isAllowed=y), [Consultado 9 de Enero 2012].
- Archambault, É., y Larivière, V. (2009). "History of the journal impact factor: Contingencies and consequences". *Scientometrics*, 79(3), 635-649.
- Arriola-Quiroz, I., Curioso, W.H., Cruz-Encarnacion, M. y Gayoso, O. (2010), "Characteristics and publication patterns of theses from a Peruvian medical school", *Health Information y Libraries Journal*, Vol. 27, No. 2, pp. 148-154.
- Bailey, B.J. (2002), "Duplicate publication in the field of otolaryngology-head and neck surgery", *Otolaryngology--head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery*, Vol. 126, No. 3, pp. 211-216.
- Barnes, T., Moyle, M., Brown, J. y Sadler, K. (2012). "Electronic doctoral theses in the UK: a sector-wide survey into policies, practice and barriers to Open Access". Disponible en: <http://discovery.ucl.ac.uk/1339905/1/UKCGE-Electronic%20Doctoral%20Theses%20in%20the%20UK%20Book.pdf>, [Consultado el 20 November 2013].
- Baufreton, C., Chrétien, J., Moreau-Cordier, F., Moreau, F., Portefaix, H., Branchereau, H., Huez, J., Richard, I. y Saint-André, J. (2012), "La production scientifique issue de la formation initiale à la faculté de médecine d'Angers entre 2002 et 2008: de bonne qualité mais insuffisante", *La Presse Médicale*, Vol. 41, No. 5, pp. e213-e219.

- Benotmane, I., Glatz, N., Bihan, S., Legrand, F., Gosset, D. y Boulanger, E. (2012), "Publications des thèses d'exercice soutenues à la faculté de médecine de Lille", *La Presse Médicale*, Vol. 41, No. 7, pp. e397-e403.
- Bergstrom, C.T. y West, J.D. (2008), "Assessing citations with the Eigenfactor metrics", *Neurology*, Vol. 71, No. 23, pp. 1850-1851.
- Bergen Communiqué, B. (2005), "The European Higher Education Area. Achieving the goals", *Communiqué of the Conference of Ministers responsible for higher education in Bergen, 19-20 May 2005*. Disponible en: [http://www.ehea.info/Uploads/Declarations/Bergen\\_Communique1.pdf](http://www.ehea.info/Uploads/Declarations/Bergen_Communique1.pdf), [Consultado el 5 de Febrero de 2013].
- Berlin Communiqué (2003), "Realising the European higher education area", *Communiqué of the Conference of Ministers responsible for higher education in Berlin on 19 September 2003*. Disponible en: [http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/mdc/berlin\\_communique1.pdf](http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/mdc/berlin_communique1.pdf), [Consultado el 2 de Febrero de 2013].
- Bethesda Statement (2003), *Bethesda statement on Open Access publishing*. Disponible en: <http://legacy.earlham.edu/~peters/fos/bethesda.htm>, [Consultado el 14 Julio de 2014].
- Bollen, J., Van de Sompel, H., Hagberg, A., y Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS One*, 4(6), e6022.
- Bologna Declaration (1999), "Towards the European Higher Education Area. Conference of Ministers responsible for Higher Education in 29 European countries (June), Bologna, Italy". Disponible en: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:c11088>, [Consultado el 8 Noviembre de 2011].
- Bordons, M., Gomez, I., Fernandez, M.T., Zulueta, M.A. y Méndez, A. (1996). "Local, domestic and international scientific collaboration in biomedical research", *Scientometrics*, Vol. 37, No. 2, pp. 279-295.
- Bordons, M. (1999). Evaluación de la actividad científica a través de indicadores bibliométricos. *Revista española de cardiología*, Vol. 52, No. 10, pp. 790-800.
- Bordons, M. y Gómez, I. (2000). "Collaboration networks in science", *The web of knowledge: A festschrift in honor of Eugene Garfield*, pp. 197-213.
- Bordons, M., Fernández, M. y Gómez, I. (2002). "Advantages and limitations in the use of impact factor measures for the assessment of research performance", *Scientometrics*, Vol. 53, No. 2, pp. 195-206.
- Borgman, C. L., y Furner, J. (2002), "Scholarly communication and bibliometrics", *Annual Review of Information Science and Technology*, Vol. 26, pp. 3-72.
- Bosman, J., Mourik, I.V., Rasch, M., Sieverts, E. y Verhoeff, H. (2006). Scopus reviewed and compared: The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar. Utrecht University Library, Utrecht.

- Bosquet, C. y Combes, P. (2013). "Are academics who publish more also more cited? Individual determinants of publication and citation records", *Scientometrics*, Vol. 97, No. 3, pp. 831-857.
- Boyack, K.W. y Klavans, R. (2010). "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?", *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 12, pp. 2389-2404.
- Bouville, M. (2008). "Plagiarism : Words or ideas" *Science and Engineering Ethics*, Vol. 14, pp. 311-322.
- Brembs, B., Button, K. y Munafò, M. (2013), "Deep impact: unintended consequences of journal rank", *Frontiers in human Neuroscience*, Vol. 7, No. 291.
- Brown, J. (2010), *Literature review of research into attitudes towards electronic theses and dissertations (ETDs)*. Disponible en: <http://discovery.ucl.ac.uk/20424/1/20424.pdf>, [Consultado el 3 de Agosto de 2013].
- Caan, W. y Cole, M. (2012), "How much doctoral research on clinical topics is published?", *Evidence-based medicine*, Vol. 17, No. 3, pp. 71-74.
- Campanario, J. y Molina, A. (2009), "Surviving bad times: The role of citations, self-citations and numbers of citable items in recovery of the journal impact factor after at least four years of continuous decreases", *Scientometrics*, Vol. 81, No. 3, pp. 859-864.
- Case, D. O., y Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635-645.
- Christianson, B., Elliot, M. y Massey, B. (2015), *The Role of Publications and Other Artefacts in Submissions for the UK PhD*. Disponible en: <https://www.ukcge.ac.uk/pdf/2015%20UKCGE-The%20Role%20of%20Publications%20and%20Other%20Artefacts%20Book%20A5%20final.pdf>, [Consultado el 12 de Septiembre de 2015].
- Cole, S., y Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American sociological review*, Vol. 32, No. 3, pp. 377-390.
- Copeland, S., Penman, A. y Milne, R. (2005), "Electronic theses: the turning point", *Program*, Vol. 39, No. 3, pp. 185-197.
- Costas, R. y Bordons, M. (2005), "Bibliometric indicators at the micro-level: some results in the area of natural resources at the Spanish CSIC", *Research evaluation*, Vol. 14, No. 2, pp. 110-120.
- Costas, R., Van Leeuwen, T.N. y Bordons, M. (2010), "Self-citations at the meso and individual levels: effects of different calculation methods", *Scientometrics*, Vol. 82, No. 3, pp. 517-537.
- Costas, R. y Bordons, M. (2011), "Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective", *Scientometrics*, Vol. 88, No. 1, pp. 145-161.

- Creative Commons. (2008), *Creative commons: Attribution-ShareAlike 3.0. United States*. Disponible en: <https://creativecommons.org/licenses/by-sa/3.0/us/legalcode> [Consultado el 8 de Abril 2015]
- Cronin, B. (2005), *The hand of science: academic writing and its rewards*. Scarecrow Press, Lanham.
- Cullen, R. y Chawner, B. (2011), "Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms", *The Journal of Academic Librarianship*, Vol. 37, No. 6, pp. 460-470.
- Cursiefen, C. y Altunbas, A. (1998), "Contribution of medical student research to the Medline™-indexed publications of a German medical faculty", *Medical education*, Vol. 32, No. 4, pp. 439-440.
- Dalton, J.T. y Seamans, N.H. (2001), *Electronic Theses and Dissertations: Two Surveys of Editors and Publishers*. Disponible en: [http://scholarworks.gsu.edu/univ\\_lib\\_facpub/47/](http://scholarworks.gsu.edu/univ_lib_facpub/47/), [Consultado el 10 de Noviembre de 2013].
- Davies, R.E. y Rolfe, G. (2009), "PhD by publication: A prospective as well as retrospective award? Some subversive thoughts", *Nurse education today*, Vol. 29, No. 6, pp. 590-594.
- Delgado-López-Cózar, E. y Cabezas-Clavijo, Á (2012), "Google Scholar Metrics: una herramienta poco fiable para la evaluación de revistas científicas", *El profesional de la información*, Vol. 21, No. 4, pp. 4207.
- Delgado-López-Cózar, E. y Repiso-Caballero, R. (2013), "El impacto de las revistas de comunicación: comparando Google Scholar Metrics, Web of Science y Scopus", *Comunicar: Revista Científica de Comunicación y Educación*, Vol. 21, No. 41, pp. 45-52.
- Dhaliwal, U., Singh, N. y Bhatia, A. (2010), "Masters theses from a university medical college: publication in indexed scientific journals", *Indian journal of ophthalmology*, Vol. 58, No. 2, pp. 101-104.
- Diez, C., Arkenau, C. y Meyer-Wentrup, F. (2000), "The German medical dissertation—Time to change?", *Academic Medicine*, Vol. 75, No. 8, pp. 861-863.
- Dinham, S. y Scott, C. (2001), "The experience of disseminating the results of doctoral research", *Journal of Further and Higher Education*, Vol. 25, No. 1, pp. 45-55.
- Echeverría, M., Stuart, D. y Blanke, T. (2015a). "Medical theses and derivative articles: dissemination of contents and publication patterns". *Scientometrics*, 102(1), 559-586.
- Echeverría, M., Stuart, D. y Blanke, T. (2015b), "Potential Predictability of References in the Identification of Derivative Articles from Doctoral Theses", *10th International Caliber 2015. March 12-14, 2015 Shimla HP India*, INFLIBNET Centre, Shimla, India, 2015, March 12-14, pp. 171-182.
- Errami, M., Hicks, J.M., Fisher, W., Trusty, D., Wren, J.D., Long, T.C. y Garner, H.R. (2008), "Deja vu--a study of duplicate citations in Medline", *Bioinformatics (Oxford, England)*, Vol. 24, No. 2, pp. 243-249.

- España. 2007. Real Decreto 1393/2007 de 29 de Octubre de 2007 por el que se establece la ordenación de las enseñanzas universitarias oficiales. BOE, Vol. 260, pp. 44037-44038
- España. 2011. Real Decreto 99/2011 de 28 de Enero de 2011 por el que se regulan las enseñanzas oficiales de doctorado. BOE, Núm. 35, pp. 13909-13926.
- European Council (2000), "Lisbon European Council (23-24 March 2000). Presidency and conclusions". Disponible en: [http://www.europarl.europa.eu/summits/lis1\\_en.htm](http://www.europarl.europa.eu/summits/lis1_en.htm), (consultado el 24 de Noviembre de 2013).
- European University Association (2005a). "Doctoral programmes for the European knowledge society, Salzburg, 3-5 February 2005. Conclusions and recommendations". Disponible en: [http://www.eua.be/Libraries/cde-website/Salzburg\\_Conclusions.pdf?sfvrsn=0](http://www.eua.be/Libraries/cde-website/Salzburg_Conclusions.pdf?sfvrsn=0), [Consultado el 3 de Marzo de 2013].
- European University Association (2005b). "Doctoral programmes for the European Knowledge Society. Final Report. EUA". Disponible en: [http://www.eua.be/eua/jsp/en/upload/Doctoral\\_Programmes\\_Project\\_Report.1129285328581.pdf](http://www.eua.be/eua/jsp/en/upload/Doctoral_Programmes_Project_Report.1129285328581.pdf), (Consultado el 10 de Octubre de 2015).
- Eurostat (2015). "Graduations in ISCED 3 to 6 by field of education and sex. Doctor of Philosophy (Ph.D), 15/9/2015 last update". Disponible en: [http://ec.europa.eu/eurostat/data/database?node\\_code=educ\\_grad5](http://ec.europa.eu/eurostat/data/database?node_code=educ_grad5), [select *Data explorer*, *Select Data*, *ISCED97*, *ED6PHP*], (Consultado el 6 Octubre de 2015).
- Falagas, M.E., Pitsouni, E.I., Malietzis, G.A. y Pappas, G. (2008), "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses", *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, Vol. 22, No. 2, pp. 338-342.
- Fawcett, T. (2006). "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27, No. 8, pp. 861-874.
- Figueredo, E., Sánchez Perales, G., Villalonga, A. y Castillo, J. (2002), "Tesis doctorales españolas sobre Anestesiología y publicaciones científicas de sus autores", *Revista Española de Anestesiología y Reanimación*, Vol. 49, No. 3, pp. 124-130.
- Fine, M.A. y Kurdek, L.A. (1993), "Reflections on determining authorship credit and authorship order on faculty-student collaborations.", *American Psychologist*, Vol. 48, No. 11, pp. 1141.
- Franceschet, M. (2010), "Journal influence factors", *Journal of Informetrics*, Vol. 4, No. 3, pp. 239-248.
- Frkovic, V., Skender, T., Dojcinovic, B. y Bilic-Zulle, L. (2002), "Publishing scientific papers based on Master's and Ph. D. theses from a small scientific community: case study of Croatian medical schools", *Croatian medical journal*, Vol. 44, No. 1, pp. 107-111.
- Gadd, E., Loddington, S. y Oppenheim, C. (2007), "A comparison of academics' attitudes towards the rights protection of their research and teaching materials", *Journal of Information Science*, Vol. 33, No. 6, pp. 686-701.
- García-Romero, A. y Estrada-Lorenzo, J.M. (2014), "A bibliometric analysis of plagiarism and self-plagiarism through Déjà vu", *Scientometrics*, Vol. 101, No. 1, pp. 381-396.

- Garfield, E., Sher, I. H. (1963), "New factors in evaluation of scientific literature through citation indexing", *American Documentation*, Vol. 14, No. 3, pp. 195–201
- Garfield, E. (1996), "When to cite", *The Library Quarterly*, Vol. 66, No. 4, pp. 449-458.
- Garfield, E. (1998), "Long-term vs. short-term journal impact: does it matter", *Scientist*, Vol. 12, No. 3, pp. 11-12.
- Garfield, E. (2000), "The use of JCR and JPI in measuring short and long term journal impact", *Croatian Medical Journal*, Vol. 41, No. 4, pp. 368-374.
- Gipp, B. y Beel, J. (2010), "Citation based plagiarism detection: a new approach to identify plagiarized work language independently". In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*. ACM, 2010, pp. 273-274.
- Gipp, B., Meuschke, N. y Beel, J. (2011), "Comparative evaluation of text-and citation-based plagiarism detection approaches using GUTTENPLAG". In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*. ACM, 2011. p. 255-258.
- Gipp, B. (2013). "Citation-based Plagiarism Detection: applying citation pattern analysis to identify currently non-machine-detectible disguised plagiarism in scientific publications". Magdeburg, Universität, Diss., 2013.
- Gipp, B. (2014). "Citation-based Plagiarism Detection". En: *Citation-based plagiarism detection*. Springer Fachmedien Wiesbaden, pp. 57-88.
- Glänzel, W. y Thijs, B. (2006), "The influence of author self-citations on bibliometric meso-indicators. The case of universities", *Scientometrics*, Vol. 66, No. 1, pp. 71-80.
- Glänzel, W. (2001), "National characteristics in international scientific co-authorship relations", *Scientometrics*, Vol. 51, No. 1, pp. 69-115.
- Glenisson, P., Glänzel, W. y Persson, O. (2005), "Combining full-text analysis and bibliometric indicators. A pilot study", *Scientometrics*, Vol. 63, No. 1, pp. 163-180.
- Golde, C.M. (2005), "The role of the department and discipline in doctoral student attrition: Lessons from four departments", *The Journal of Higher Education*, Vol. 76, No. 6, pp. 669-700.
- Gorraiz, J. y Schloegl, C. (2008), "A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science", *Journal of Information Science*, Vol. 34, No. 5, pp. 715-725.
- Greig, M. (2005), "Implementing electronic theses at the University of Glasgow: cultural challenges", *Library collections, acquisitions, and technical services*, Vol. 29, No. 3, pp. 326-335.
- Hagen, N. (2010), "Deconstructing doctoral dissertations: how many papers does it take to make a PhD?", *Scientometrics*, Vol. 85, No. 2, pp. 567-579.
- Harnad, S. (2009), *Integrating universities' thesis and research deposit mandates*. Disponible en: <http://www.archipel.uqam.ca/6576/1/ETDpaper.pdf>. [Consultado el 12 de Junio 2014]

- Hearst, M.A. y Pedersen, J.O. (1996), "Reexamining the cluster hypothesis: scatter/gather on retrieval results", *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 76-84.
- Hoddell, S., Street, D. y Wildblood, H. (2002), "Doctorates-converging or diverging patterns of provision", *Quality Assurance in Education*, Vol. 10, No. 2, pp. 61-70.
- Hollmann, M., Borrell, C., Garin, O., Fernández, E. y Alonso, J. (2015), "Factors influencing publication of scientific articles derived from masters theses in public health", *International journal of public health*, Vol. 60, No. 4, pp. 495-504.
- International Committee of Medical Journal Editors (2010), *Uniform requirements for manuscripts submitted to biomedical journals*. Disponible: <http://www.icmje.org>, [Consultado el 2 de Abril 2015].
- Introna, L. y Hayes, N. (2004), "Plagiarism, detection and intentionality: on the construction of plagiarists", *Plagiarism: Prevention, Practice and Policies, 28th June-30th June*, .
- Iribarren Maestro, I. (2006), *Producción científica y visibilidad de los investigadores de la Universidad Carlos III de Madrid en las bases de datos del ISI, 1997-2003*. Tesis doctoral. Disponible en: <http://e-archivo.uc3m.es/bitstream/handle/10016/1088/TESIS%20IIM.pdf;jsessionid=5E57706A21B2D261C4D494EDEC1EB36?sequence=1>. [Consultado el 12 de Marzo de 2014].
- ISO (1986) (1986), "International Standard 7144, presentation of theses and similar documents", *International Organization for Standardization, Documentation and Information*, 3rd ed. ed., ISO, Gèneve, pp. 679-688.
- Janssens, F., Tran Quoc, V., Glänzel, W. y De Moor, B. (2006), "Integration of textual content and link information for accurate clustering of science fields" En: *Proceedings of the I International Conference on Multidisciplinary Information Sciences y Technologies (InSciT2006). Current Research in Information Sciences and Technologies. Volume I*, pp. 615.
- Johnston, D.W., Piatti, M. y Torgler, B. (2013), "Citation success over time: theory or empirics?", *Scientometrics*, Vol. 95, No. 3, pp. 1023-1029.
- Jones, R. y Andrew, T. (2005), "Open access, open source and e-theses: the development of the Edinburgh Research Archive", *Program*, Vol. 39, No. 3, pp. 198-212.
- Jones, R.E., Andrew, T. y MacColl, J. (2006), *The institutional repository*, Oxford, Chandos.
- Kamler, B. (2008), "Rethinking doctoral publication practices: Writing from and beyond the thesis", *Studies in Higher Education*, Vol. 33, No. 3, pp. 283-294.
- Katz, J.S. y Martin, B.R. (1997), "What is research collaboration?", *Research policy*, Vol. 26, No. 1, pp. 1-18.
- Kessler, M.M. (1963) "Bibliographic coupling between scientific papers", *American Documentation*, Vol. 14, No. 1, pp. 10-25.
- Kumar, R. and Indrayan, A. (2011), "Receiver operating characteristic (ROC) curve for medical researchers", *Indian pediatrics*, Vol. 48, No. 4, pp. 277-287.

- Kyvik, S. y Smeby, J. (1994), "Teaching and research. The relationship between the supervision of graduate students and faculty research performance", *Higher Education*, Vol. 28, No. 2, pp. 227-239.
- Larivière, V., Zuccala, A. y Archambault, É (2007), "The declining scientific impact of theses: Implications for electronic thesis and dissertation repositories and graduate studies", *Scientometrics*, Vol. 74, No. 1, pp. 109-121
- Larivière, V. (2010), *A bibliometric analysis of Quebec's PhD students' contribution to the advancement of knowledge*. Thesis unpublished. McGill University.  
 Disponible en:  
[http://www.ost.uqam.ca/Portals/0/docs/Monographies/Thesis\\_Lariviere\\_Final.pdf](http://www.ost.uqam.ca/Portals/0/docs/Monographies/Thesis_Lariviere_Final.pdf),  
 [Consultada el 4 de Diciembre de 2014].
- Lariviere, V. y Gingras, Y. (2010), "On the prevalence and scientific impact of duplicate publications in different scientific fields (1980-2007)", *Journal of Documentation*, Vol. 66, No. 2, pp. 179-190.
- Larivière, V. (2011), "On the shoulders of students? The contribution of PhD students to the advancement of knowledge", *Scientometrics*, Vol. 90, No. 2, pp. 463-481.
- Leash, E. (1997), "Is it time for a new approach to authorship?", *Journal of dental research*, Vol. 76, pp. 724-727.
- López-Illescas, C., Moya-Anegón, F. y Moed, H.F. (2009), "Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus:: the effect of poorly cited journals in oncology", *Journal of Information Science*, Vol. 35, No. 2, pp. 244-256.
- Lozano, G. A., Larivière, V., y Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 11, pp. 2140-2145.
- Macduff, C. (2009), "An evaluation of the process and initial impact of disseminating a nursing e-thesis", *Journal of advanced nursing*, Vol. 65, No. 5, pp. 1010-1018.
- Martin, T., Ball, B., Karrer, B. y Newman, M. (2013), "Coauthorship and citation in scientific publishing", *arXiv preprint arXiv:1304.0473*, Disponible en:  
<http://arxiv.org/abs/1304.0473>. [Consultado el 5 Mayo de 2014].
- Mattsson, P., Sundberg, C.J. y Laget, P. (2010), "Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position", *Scientometrics*, Vol. 87, No. 1, pp. 99-105.
- McMillan, G., Ramirez, M.L., Dalton, J., Read, M. y Seamans, N.H. (2011), "An Investigation of ETDs as Prior Publications: Findings from the 2011 NDLTD Publishers' Survey".  
 Disponible en: <http://hdl.handle.net/10919/11338> [Consultado el 20 de Septiembre de 2014]
- Merton, R.K. (1968), "The Matthew effect in science", *Science*, Vol. 159, No. 3810, pp. 56-63.
- Merton, R.K. (1973), *The sociology of science: Theoretical and empirical investigations* University of Chicago press.

- Moed, H.F. (2002), "The impact-factors debate: the ISI's uses and limits", *Nature*, Vol. 415, No. 6873, pp. 731-732.
- Nettles, M. T. y Millett, C. M. (2006), *Three magic letters: getting to PH.D.* Johns Hopkins University Press, Baltimore, Md.
- Nieminen, P., Sipila, K., Takkinen, H.M., Renko, M. y Risteli, L. (2007), "Medical theses as part of the scientific training in basic medical and dental education: experiences from Finland", *BMC medical education*, Vol. 7, pp. 51.
- Observatoire des Sciences and Techniques. (2002), "Indicateurs bibliometriques des institutions publiques de recherche.". Mentioned by Paillassard, P., Schöpfel., Stock, C. (2007) (Ed.), *Dissemination and preservation of French print and electronic theses. The Grey Journal*, 3(2), pp. 77-93.
- Paillassard, P., Schöpfel, J. y Stock, C. (2007), "Dissemination and preservation of French print and electronic theses", *The Grey Journal*, Vol. 3, No. 2, pp. 77-93.
- Park, C. (2007), *Redefining the doctorate*, The Higher Education Academy. Disponible en: [https://www.heacademy.ac.uk/sites/default/files/redefining\\_the\\_doctorate.pdf](https://www.heacademy.ac.uk/sites/default/files/redefining_the_doctorate.pdf), [Consultado el 20 de Octubre de 2013].
- Pertile, S.d.L., Rosso, P. y Moreira, V.P. (2013), "Counting Co-occurrences in Citations to Identify Plagiarised Text Fragments", *Information Access Evaluation. Multilinguality, Multimodality, and Visualization* Springer, pp. 150-154.
- Pertile, S.d.L., Moreira, V.P. y Rosso, P. (2015), "Comparing and combining Content-and Citation-based approaches for plagiarism detection", *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.23593. Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23593>. [Consultado 9 de Septiembre de 2015]
- Pitche, P.T., Onipoh, D. y Tchangai-Walla, K.L. (2007), "Devenir scientifique des thèses pour le diplôme d'état de doctorat en médecine soutenues à l'université de Lomé (Togo)", *Pédagogie médicale*, Vol. 8, No. 1, pp. 24-29.
- Pole, C.J., Sprokkereef, A., Burgess, R.G. y Lakin, E. (1997), "Supervision of doctoral students in the natural sciences: Expectations and experiences", *Assessment y Evaluation in Higher Education*, Vol. 22, No. 1, pp. 49-63.
- Powell, S. (2004), *The award of PhD by published work in the UK*, UK Council for Graduate Education, Lichfield. Disponible en: [http://www.ukcge.ac.uk/pdf/PhD%20by%20Published%20Work%20in%20the%20UK%20\(2004\).pdf](http://www.ukcge.ac.uk/pdf/PhD%20by%20Published%20Work%20in%20the%20UK%20(2004).pdf), [Consultado el 20 de Noviembre de 2013].
- Quality Assurance Agency for Higher Education (2011), *Doctoral degree characteristics*. Disponible en: <http://www.qaa.ac.uk>, [Consultado el 4 de Mayo de 2012].
- Rennie, D., Yank, V. y Emanuel, L. (1997), "When authorship fails: a proposal to make contributors accountable", *JAMA*, Vol. 278, No. 7, pp. 579-585.
- Salmi, L.R., Gana, S. y Mouillet, E. (2001), "Publication pattern of medical theses, France, 1993–98", *Medical education*, Vol. 35, No. 1, pp. 18-21.

- Sanz Casado, E., y Moreno, C. M. (1997). Técnicas bibliométricas aplicadas a los estudios de usuarios. *Revista general de información y documentación*, Vol.7, No.2, pp. 41-68
- Schein, M., y Paladugu, R. (2001). Redundant surgical publications: tip of the iceberg?. *Surgery*, 129(6), 655-661.
- Schubert, A., Glänzel, W. y Thijs, B. (2006), "The weight of author self-citations. A fractional approach to self-citation counting", *Scientometrics*, Vol. 67, No. 3, pp. 503-514.
- Seamans, N.H. (2003), "Electronic theses and dissertations as prior publications: what the editors say", *Library Hi Tech*, Vol. 21, No. 1, pp. 56-61.
- Seglen, P.O. (1997), "Why the impact factor of journals should not be used for evaluating research", *BMJ*, Vol. 314, No. 7079, pp. 497.
- Shapiro, D. W., Wenger, N. S., y Shapiro, M. F. (1994). The contributions of authors to multiauthored biomedical research papers. *JAMA*, Vol. 271, No. 6, pp. 438-442.
- Sharmini, S., Spronken-Smith, R., Golding, C. y Harland, T. (2015), "Assessing the doctoral thesis when it includes published work", *Assessment y Evaluation in Higher Education*, Vol. 40, No. 1, pp. 89-102.
- Shibayama, S. y Baba, Y. (2015), "Impact-oriented science policies and scientific publication practices: The case of life sciences in Japan", *Research Policy*, Vol. 44, No. 4, pp. 936-950.
- Sollaci, L.B. y Pereira, M.G. (2004), "The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey", *Journal of the Medical Library Association : JMLA*, Vol. 92, No. 3, pp. 364-367.
- Sorokina, D., Gehrke, J., Warner, S. y Ginsparg, P. (2006), "Plagiarism detection in arXiv", "Plagiarism detection in arXiv", *Data Mining, 2006. ICDM'06*. En: *Sixth International Conference on IEEE* (2006) pp. 1070-1075.
- Stanton, K.V. y Liew, C.L. (2011), "Risks, Benefits and Revelations: An Exploratory Study of Doctoral Students' Perceptions of Open Access Theses in Institutional Repositories", En: *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation* Springer, Vol. 7008 de la serie "Lecture Notes on Computer Science" pp. 182-191.
- Suber, P. (2008), "Open access to electronic theses and dissertations", *DESIDOC Journal of Library y Information Technology*, Vol. 28, No. 1, pp. 25-34.
- Subramanyam, K. (1983), "Bibliometric studies of research collaboration: A review", *Journal of Information Science*, Vol. 6, No. 1, pp. 33-38.
- Sun, Z., Errami, M., Long, T., Renard, C., Choradia, N. y Garner, H. (2010), "Systematic characterizations of text similarity in full text biomedical publications", *PloS One*, Vol. 5, No. 9, pp. e12704.
- Sursock, A., Smidt, H. y Davies, H. (2010), *Trends 2010: a decade of change in European Higher Education*. Vol. 1, Brussels, European University Association. Disponible en: [http://collectif-papera.org/IMG/pdf/EUA\\_Trends\\_2010-2.pdf](http://collectif-papera.org/IMG/pdf/EUA_Trends_2010-2.pdf), [Consultado el 12 de Septiembre de 2015].

- Thelwall, M., Vaughan, L., y Björneborn, L. (2005), "Webometrics", in B. Cronin(Ed.), *Annual Review of Information Science and Technology*. 39 Information Today, Medford,NJ, pp. 81-135.
- Thelwall, M. (2008), "Bibliometrics to webometrics", *Journal of Information Science*, Vol. 34, No. 4, pp. 1-18.
- Thompson, L.A. (2001), "Electronic theses and dissertations at Virginia Tech", *Science y Technology Libraries*, Vol. 20, No. 1, pp. 87-101.
- Thomson Reuters. (2012), *The Thomson Reuters journal selection process*. Updated 5-2012 by Jim Testa. Disponible en: <http://wokinfo.com/essays/journal-selection-process>. [Consultado el 2 Octubre de 2015].
- Torres-Salinas, D. (2007), "Diseño de un sistema de información y evaluación científica. Análisis cuantitativo de la actividad investigadora de la Universidad de Navarra en el área de Ciencias de la Salud. 1999-2005". Thesis unpublished. Universidad de Granada. Disponible en: <http://eprints.rclis.org/10545/>, [Consultado el 4 de Junio de 2013].
- Torres-Salinas, D., Lopez-Cózar, E.D. y Jiménez-Contreras, E. (2009), "Ranking of departments and researchers within a university using two different databases: Web of Science versus Scopus", *Scientometrics*, Vol. 80, No. 3, pp. 761-774.
- Troman, A., Jacobs, N. y Copeland, S. (2007), "A new electronic service for UK theses: access transformed by EThOS", *Interlending y Document Supply*, Vol. 35, No. 3, pp. 157-163.
- Tscharntke, T., Hochberg, M.E., Rand, T.A., Resh, V.H. y Krauss, J. (2007), "Author sequence and credit for contributions in multiauthored publications", *PLoS Biol*, Vol. 5, No. 1, pp. e18.
- UK Copyright Service (2012), *Copyright issues for derivative works: Fact sheet P-22*. Disponible en: [http://www.copyrightservice.co.uk/copyright/p22\\_derivative\\_works](http://www.copyrightservice.co.uk/copyright/p22_derivative_works), [Consultado el 12 de Mayo de 2014].
- van Eyk, H.J., Hooiveld, M.H., Van Leeuwen, T.N., Van der Wurff, Bert LJ, De Craen, A.J. and Dekker, F.W. (2010), "Scientific output of Dutch medical students", *Medical teacher*, Vol. 32, No. 3, pp. 231-235.
- von Elm, E., Poglia, G., Walder, B., y Tramer, M. R. (2004). Different patterns of duplicate publication: an analysis of articles used in systematic reviews. *JAMA*, 291(8), 974-980.
- Wang, J. (2013), "Citation time window choice for research impact evaluation", *Scientometrics*, Vol. 94, No. 3, pp. 851-872.
- Watts, J.H. (2012), "To publish or not to publish before submission? Considerations for doctoral students and supervisors", *Educational Practice and Theory*, Vol. 34, No. 1, pp. 65-79.
- Weeks, W.B., Wallace, A.E. y Kimberly, B.S. (2004), "Changes in authorship patterns in prestigious US medical journals", *Social science y medicine*, Vol. 59, No. 9, pp. 1949-1954.

- Weiss Sholom M., Indurkha, N. y Zhang, T. (2010), "Overview of text mining", in S.M. Weiss(Ed.), *Fundamentals of predictive text mining* Springer, London, pp. 1-12.
- Wouters, P.F. (1999), *The citation culture*. Unpublished thesis. University of Amsterdam.  
Disponible en: <http://dare.uva.nl/document/2/8218>, [Consultado el 25 de Septiembre de 2015].
- Zamudio Igami, M., Bressiani, J.C. y Mugnaini, R. (2014), "A new model to identify the productivity of theses in terms of articles using co-word analysis", *Journal of Scientometric Research*, Vol. 3, No. 1, pp. 3.
- Ziemann, E. y Oestmann, J. (2012), "Publications by doctoral candidates at charité university hospital, Berlin, from 1998–2008", *Deutsches Ärzteblatt International*, Vol. 109, No. 18, pp. 333.
- Zuckerman, H.A. (1968), "Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity", *American Journal of Sociology*, , pp. 276-291.