



VNiVERSiDAD
D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL



UNIVERSIDAD DE SALAMANCA FACULTAD DE
TRADUCCIÓN Y DOCUMENTACIÓN
MÁSTER EN SISTEMAS DE INFORMACIÓN DIGITAL
Trabajo Fin de Máster

Etiquetado de información digital

Análisis y valoración de su aplicación en los Trabajos
de Fin de Máster del Repositorio GREDOS

Realizado por Laura Moreno López
y dirigido por Crispulo Travieso Rodríguez

Salamanca, 2017



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

MÁSTER EN SISTEMAS DE
INFORMACIÓN DIGITAL

UNIVERSIDAD DE SALAMANCA FACULTAD DE
TRADUCCIÓN Y DOCUMENTACIÓN
MÁSTER EN SISTEMAS DE INFORMACIÓN DIGITAL
Trabajo Fin de Máster

Etiquetado de información digital

Análisis y valoración de su aplicación en los Trabajos
de Fin de Máster del Repositorio GREDOS

Realizado por Laura Moreno López
TUTOR: Crispulo Travieso Rodríguez

Salamanca, 2017

ASIENTO CATALOGRÁFICO ADAPTADO AL REPOSITORIO INSTITUCIONAL GREDOS

Título: Etiquetado de información digital: análisis y valoración de su aplicación en los Trabajos de Fin de Máster del Repositorio GREDOS.

Autor (es): Moreno López, Laura

Director(es): Travieso Rodríguez, Crispulo

Palabras clave:

[ES]: Universidad de Salamanca (España), Gredos (Repositorio), Descripción de información, Etiquetado libre, Etiquetado social, Vocabularios controlados, Lenguaje natural, Análisis terminológico

[EN]: University of Salamanca (Spain), Gredos (repository), Description information, Free tagging Social Tagging, Controlled vocabulary, Natural language, Terminological analysis

Clasificación UNESCO: 57 Lingüística: 5701 Lingüística aplicada: 570105 Lenguajes documentales

Fecha: 2017-07-17

Descripción: Trabajo de Fin de Máster en Sistemas de Información Digital, 2017

Resumen [ES]: La descripción de información ha sido una tarea llevada a cabo por profesionales de la documentación mediante una indización controlada. Con los avances en las tecnologías de la información, esta responsabilidad caerá también sobre autores y usuarios de la Web, con la asignación de etiquetas en un lenguaje natural. Este proceso de indización libre es conocido como etiquetado o “tagging” y se presenta como la alternativa para la descripción de información digital. El objetivo fundamental de este trabajo es analizar el procedimiento de asignación de etiquetas/palabras clave por parte de los autores en la descripción de TFMs ubicados en el repositorio institucional GREDOS, y reflexionar sobre si el etiquetado en lenguaje natural podría sustituir a los lenguajes documentales tradicionales. Los resultados obtenidos muestran que las palabras clave/etiquetas asignadas en lenguaje natural, aun presentan carencias que deben resolverse, ya que por el momento no alcanzan el nivel de descripción que proporcionan los vocabularios controlados, aunque estos también cuentan con sus respectivas limitaciones. Por lo tanto, actualmente la indización llevada a cabo por profesionales a partir de vocabularios controlados no debe ser reemplazada por la indización realizada por autores en lenguaje natural.

Abstract [EN]: The description of information has been a work done by librarians through controlled indexing. With advances in information technology this responsibility is also performed by authors and web users, with the allocation of tags in a natural language. This free indexing process is known as tagging and it is a possible alternative to the description of digital information. The main objective of this research is to analyze the allocation procedure of tags/keywords by the authors in their TFMs located at institutional repository GREDOS, in order to reflect on whether tagging in natural language could replace traditional documentary language. Results show that the keywords/tags assigned in natural language still have problems that must be resolved, because for now they don't get the level of description provided by controlled vocabularies. Therefore, now the professional indexing shouldn't be replaced by natural language indexing.

Tabla de contenido

1	Introducción	5
1.1	Justificación	6
1.2	Objetivos	7
2	La descripción de contenidos en entorno digitales.	8
2.1	Evolución de la descripción mediante palabras clave	8
2.2	La irrupción de la web 2.0	9
2.2.1	Etiquetado libre	10
2.2.2	Etiquetado social	12
2.2.3	Folksonomías	13
2.2.4	Lenguaje natural versus vocabularios controlados	16
2.3	La terminología en la Documentación	17
2.4	El uso de etiquetas en unidades de información	19
2.5	Descripción de recursos científicos	24
2.6	El etiquetado social y su aplicación documental: estudios previos	25
2.7	Repositorio institucional GREDOS	29
3	Metodología	32
3.1	Revisión bibliográfica	32
3.2	Estudio de caso	33
3.2.1	Variables de estudio	34
3.2.2	Población y muestra	38
3.2.3	Métodos de análisis	38
3.2.4	Limitaciones y problemas encontrados	40
4	Resultados y discusión	41
5	Conclusiones	61
	Bibliografía	65
	Anexo	1

Índice de imágenes

Imagen 1. Fases de la creación de palabras clave (Elaboración propia).....	8
Imagen 2. Nube de etiquetas del catálogo del Sistema Nacional de Bibliotecas de Euskadi.....	11
Imagen 3. Ventajas e inconvenientes de las folksonomías (Elaboración propia).....	14
Imagen 4. Interfaz del OPAC de la biblioteca digital Infobook, destacando la opción de búsqueda “Por tags”	20
Imagen 5. Registro de un recurso en Infobook, destacando la sección de tags asignadas por los propios usuarios.....	20
Imagen 6. Registro de un libro electrónico en CIELO, destacando las etiquetas asignadas por los propios usuarios.....	21
Imagen 7. Interfaz de agregación de etiquetas en CIELO.....	21
Imagen 8. Sección de etiquetado dentro de un registro bibliográfico en Worldcat.....	22
Imagen 9. Gestión personal de etiquetas en Worldcat.....	22
Imagen 10. Nube de etiquetas en un registro de AADL.	22
Imagen 11. Interfaz de un registro documental en Mendeley.	24
Imagen 12. Página inicial de GREDOS.....	30
Imagen 13. Distribución en la hoja de cálculo utilizada para el análisis de las palabras clave	39

Índice de tablas

Tabla 1. Comparativa entre tesauros y folksonomías (Soler Monreal & Gil Leiva, 2010)	16
Tabla 2. Ramas de másters con TFM con palabras clave de autor.....	42
Tabla 3. Resultados de los tipos de etiquetas simples (análisis gramatical).....	44
Tabla 4. Resultados de los tipos de etiquetas simples (análisis gramatical). Resultados obtenidos por Spiteri (2007).....	45
Tabla 5. Total de etiquetas en función del número de términos que la forman.....	46
Tabla 6. Total de etiquetas presentes en los diferentes apartados del TFM.....	56
Tabla 7. Número de etiquetas por TFM.....	59

Índice de gráficos

Gráfico 1. Distribución porcentual de ramas de máster con TFM con palabras clave..	42
Gráfico 2. Distribución porcentual de la aparición de etiquetas simples y compuestas	43
Gráfico 3. Distribución porcentual de los tipos de etiquetas simples (análisis gramatical).....	44
Gráfico 4. Distribución porcentual de la aparición de etiquetas en otro idioma diferente al castellano.....	48
Gráfico 5. Distribución porcentual de las etiquetas neologismo	49
Gráfico 6. Distribución porcentual de las etiquetas con y sin errores ortográficos	50
Gráfico 7. Distribución porcentual de la aparición de etiquetas en plural y singular	51
Gráfico 8. Distribución porcentual de las etiquetas que coinciden en el texto de sus respectivos TFMs	52
Gráfico 9. Distribución porcentual de las etiquetas con y sin sinónimos dentro del TFM	53
Gráfico 10. Distribución porcentual de las etiquetas con o sin variación morfológica dentro del TFM.	53

Gráfico 11. Distribución porcentual de las etiquetas en función de su aparición en el texto. Resultados obtenidos por Heckner et al. (2008).....	55
Gráfico 12. Distribución porcentual de las etiquetas en función del contenido que representan	57
Gráfico 13. Distribución porcentual de la coincidencia y no coincidencia de las etiquetas en GREDOS.....	58

1 Introducción

Desde la irrupción de Internet, se ha producido una evolución constante dando lugar a un nuevo espacio donde la generación de conocimiento y el poder y actuación del usuario han cambiado por completo. Con la llegada de la Web 2.0 se consigue derribar la barrera entre consumidor y creador de información. Este nuevo perfil proactivo de los usuarios ha dado lugar a una sociedad de la información donde cada vez es mayor la creación, distribución y acceso a recursos digitales de una forma descentralizada y distribuida.

Ante este nuevo paradigma surge la necesidad de encontrar soluciones frente al crecimiento exponencial de la información digital, o, lo que es lo mismo, a una sobrecarga informativa o “sobreinformación”. En un primer momento, los profesionales de la documentación han sido los responsables de llevar a cabo la descripción de los recursos mediante una indización por medio de vocabularios controlados, donde seleccionaban descriptores que mejor describían o representaban la información para su futura recuperación, es decir, las palabras clave han sido y son una herramienta de descripción de contenidos básica. Sin embargo, con la Web 2.0 esta responsabilidad caerá también sobre autores y usuarios, mediante la asignación de etiquetas en un lenguaje natural. Este proceso de indización libre es conocido como etiquetado o “tagging” y se presenta como la alternativa para la descripción de información digital.

Gracias a la descripción de información mediante un lenguaje natural, es decir, por medio del etiquetado libre, el autor puede describir por sí mismo el contenido que crea o que consume. Por esta razón es de gran importancia comprender cómo surge, evoluciona y se realiza este proceso y cuáles son sus ventajas e inconvenientes en la descripción de información digital, ya que, como se expone, existe una gran controversia respecto a los beneficios y problemas del etiquetado libre. Por lo tanto, la razón de ser de este trabajo es conocer el comportamiento del autor en el proceso de asignación de etiquetas, es decir, analizar el modo o las pautas de etiquetado de contenidos. Para ello, en la primera parte del trabajo se ha realizado una revisión bibliográfica sobre esta materia, para conocer y ubicar el estado de la cuestión, y a continuación se ha llevado a cabo un estudio de caso donde se han establecido unos criterios de análisis que han sido aplicados a un total de 963 palabras clave asignadas por los autores (en lenguaje natural) de los Trabajos de Fin de Máster¹, ubicados en el Repositorio GREDOS de la Universidad de Salamanca². Se ha escogido este repositorio ya que al realizar las prácticas académicas en dicho centro se ha podido observar el etiquetado libre desde una perspectiva profesional en un entorno donde se gestiona documentación digital, además de presentarse como un escenario ideal para el análisis de etiquetas asignadas por sus propios autores.

Con este trabajo se pretende dar visibilidad a la importancia que tiene este nuevo modelo de indización, ya que aunque sea una temática con una amplia bibliografía y estudiada por la comunidad científica, son menos numerosos los estudios que analizan de una forma detallada (desde la sintaxis y lingüística) las etiquetas para poder comprender el comportamiento del usuario ante este proceso.

Finalmente, tras el análisis de los resultados cabe reflexionar sobre si la descripción de contenidos llevado a cabo por los autores mediante el etiquetado libre está siendo implementada de forma idónea. En caso afirmativo, ¿dejaría atrás este nuevo modelo de indización libre a los vocabularios controlados tradicionales?

¹ En adelante TFM (en plural TFMs).

² En adelante USAL.

1.1 Justificación

Este trabajo se presenta como culminación del Máster en Sistemas de Información Digital de la Universidad de Salamanca en el curso 2016-2017. Para su realización fue esencial el Módulo II, llamado Producción, procesamiento y recuperación de la información digital, con la asignatura “Descripción, representación y organización de contenidos digitales”, donde se pudo conocer la organización de los contenidos digitales en repositorios y bibliotecas digitales, junto con la creación de mapas conceptuales y las características y tipos de la indización social frente al uso de vocabularios controlados. Además, este trabajo está relacionado con las prácticas académicas realizadas durante el máster en el Repositorio Documental de la Universidad de Salamanca GREDOS. GREDOS es una plataforma de gestión de conocimiento donde se recopila, gestiona, organiza y da acceso a todos los recursos digitales producidos o custodiados por la Universidad de Salamanca.

Es importante destacar que las funciones realizadas durante el período de prácticas me han ayudado a comprender de una forma más clara el objetivo de este estudio además de observar este fenómeno desde una perspectiva y un entorno profesional. A lo largo de las 5 semanas, mi tarea principal se centró en la subida al repositorio y descripción mediante esquemas de metadatos de tesis doctorales, a través del software de gestión DSpace. Y en un menor periodo de tiempo lleve a cabo la normalización y modificación de autoridades, puesto que existían varias formas para un solo autor.

En un inicio el objetivo era analizar el modo o las pautas del etiquetado, partiendo de que las palabras clave que aparecían como materias de los TFMs en GREDOS eran las asignadas por los propios autores, pero al comparar las palabras clave de TFMs y las de su registro en GREDOS, se observó que no todas coincidían. Gracias a la realización de las prácticas, se pudieron conocer los entresijos y funcionamiento del repositorio, por lo que se llegó a la conclusión de que una de las razones por las que en algunos casos no se han respetado las palabras clave de autor, es debido a la necesidad de normalización de materias asignando términos admitidos que signifiquen lo mismo. Este es un posible inconveniente de que existan varias personas subiendo recursos al repositorio, aunque estén establecidas unas instrucciones claras. Por este motivo, aparte de analizar las palabras clave asignadas por sus autores de todos los TFMs ubicados en el repositorio, se ha comparado la coincidencia de estas etiquetas en GREDOS y en los propios TFMs.

Este TFM sigue las indicaciones establecidas por la Comisión de TFM del Máster en Sistemas de Información Digital en el curso 2016-2017 y se presenta como medio de asentamiento de los conocimientos teóricos y prácticos adquiridos y de las aptitudes alcanzadas durante el curso. Con el propósito de conseguir una comprensión del etiquetado de información digital, se presenta este estudio orientado hacia el análisis de la terminología recopilada, comportamiento del autor/usuario y utilidad de la descripción de contenidos en el contexto del repositorio GREDOS; además, pretende brindar una mayor visibilidad a este nuevo entorno como un tema relevante para su estudio.

1.2 Objetivos

El objetivo principal del trabajo es el siguiente:

Analizar el procedimiento de asignación de etiquetas/palabras clave por parte de los autores en la descripción de TFMs.

Entre los objetivos específicos encontramos los siguientes:

- Revisar la bibliografía existente sobre la utilización de palabras clave como medio de descripción de información y estudios realizados sobre el análisis de etiquetas en lenguaje natural.
- Determinar, en función de los estudios revisados, un conjunto de variables para el análisis de las palabras clave recopiladas.
- Comparar el grado de coincidencia de las palabras clave dentro del TFM, con las presentes en el registro en GREDOS.
- Diferenciar la etiqueta en función del número de términos que la forma y sus características lingüísticas y gramaticales.
- Analizar la etiqueta en función de su contenido y su ubicación dentro de las diferentes partes del TFM (título, resumen o cuerpo).
- Valorar, a la luz de los resultados obtenidos, si el etiquetado en lenguaje natural podría sustituir a los lenguajes documentales tradicionales.

A continuación se presenta una introducción teórica a este fenómeno, su origen y características, y otros estudios que comparten el mismo interés. Posteriormente, en el apartado de metodología, se desarrolla en profundidad las diferentes partes del trabajo (teórica y práctica), describiendo las técnicas de recogida y análisis de datos, y presentando los resultados. Finalmente se expondrán los resultados junto con las conclusiones finales.

2 La descripción de contenidos en entorno digitales.

2.1 Evolución de la descripción mediante palabras clave

El proceso de indización ha sido y es una de las tareas básicas para la descripción de información. El objetivo de esta actividad es representar el contenido para una adecuada gestión y futura localización de la información. Las palabras clave (*keywords*) son el resultado de este análisis textual y ofrecen información (temática, técnica o metodológica) que orienta al bibliotecario o al usuario a encontrar la información que desea, incluso a ampliar sus resultados de búsqueda.

La norma ISO 5963:1985, cuyo equivalente en español es la UNE 50-121-91, indica que para realizar un correcto análisis documental es aconsejable prestar atención a apartados relevantes del documento (título, resumen y primeros párrafos de capítulos), además de palabras o frases tipográficamente destacadas. De esta forma se conseguirán palabras clave (y futuros descriptores) suficientemente significativas que identifiquen adecuadamente el contenido del recurso y ahorren tiempo y esfuerzo en la búsqueda de información a los usuarios, por el contrario el uso incorrecto de las mismas dificultaría todo este proceso (Gil-Leiva & Alonso-Arroyo, 2005).

La descripción y representación de información mediante palabras clave comenzó siendo una tarea tradicional en la catalogación manual de fondos. Este proceso ha ido acompañado de herramientas de organización del conocimiento y control del vocabulario (tesauros, ontologías o taxonomías), que han traducido palabras clave en términos controlados y unívocos, llamados descriptores. De modo que, tras un análisis documental y posterior extracción de palabras clave, estas son cotejadas con vocabularios controlados, obteniendo el descriptor adecuado para describir el recurso. Por lo tanto, las palabras clave funcionan como un primer paso o fase inicial para una indización más desarrollada (Imagen 1).



No obstante, con la aplicación de tecnologías de tratamiento de información, empezaron a surgir nuevas modalidades automáticas o semiautomáticas de asignación de palabras clave. En un inicio, se emplean básicamente técnicas estadísticas para obtener los términos de indización. Posteriormente se incorporan algoritmos que analizan el texto, teniendo en cuenta cuestiones como la frecuencia o el lugar de aparición de palabras, y ofreciendo como resultado un conjunto de términos con cierta relevancia. El objetivo de automatizar este proceso no es más que minimizar

el coste humano y reducir y agilizar el tiempo en la tarea de la indización manual de un gran volumen de información (Motejo Ráez, 2003). A partir de la década de los ochenta se propuso incorporar técnicas de procesamiento del lenguaje natural en la indización automática, para poder alcanzar una mayor precisión. Pero finalmente, estas propuestas eran combinadas con algoritmos estadísticos. Los sistemas de indización automática siguen surgiendo y buscando soluciones para alcanzar resultados cada vez más eficientes (Gil-Leiva, 2003).

Como cabría esperar, el proceso de descripción de información ha seguido evolucionando y adaptándose a los nuevos entornos. Con el desarrollo imparable de las Tecnologías de la Información y Comunicación³, el uso de palabras clave se ha dirigido hacia la Web 2.0 y Web Semántica, donde las etiquetas (nombre que adquieren las palabras clave en el entorno digital) se convierten en una herramienta básica para describir y organizar información, creada por el propio usuario de una forma espontánea y en lenguaje natural.

En conclusión, el uso y aplicación de palabras clave ha sido y es esencial, tanto en entornos tradicionales (bibliotecas, centros de documentación y archivos) como en entornos digitales (repositorios digitales, gestores de referencias, marcadores sociales, etc.) y es innegable que las palabras clave o etiquetas, independientemente del modo en que son generadas, son un valor añadido en la gestión, localización y difusión de la información.

2.2 La irrupción de la web 2.0

En las últimas décadas la evolución de la Web junto con las TICs ha sido vertiginosa, dando lugar a un entorno web en continuo desarrollo y a un crecimiento exponencial de la información digital. Frente al inmovilismo, estaticidad y falta de comunicación de la web 1.0, nace una nueva versión donde el papel del usuario cambia por completo: esta es la denominada Web 2.0, un espacio donde el usuario pasa de ser un espectador, que básicamente buscaba y consumía información, a ser un usuario proactivo que la genera y comparte.

Estos nuevos usuarios son conscientes del esfuerzo que requiere seguir el ritmo de las tecnologías (a diferencia de los “nativos digitales”), pues este nuevo entorno digital ha modificado la sociedad ofreciendo un nuevo espacio participativo y descentralizado. Es decir, si el medio se transforma, también cambia la participación por parte de los usuarios y la accesibilidad de la información (Díaz Piraquive, Joyanes Aguilar & Medina García, 2009).

La Web 2.0, según uno de sus promotores O'Reilly (2005) se caracteriza por un conjunto de principios, de los cuales se destacan:

- Aprovechar la inteligencia colectiva de los usuarios.
- Beneficiarse de las experiencias enriquecedoras de los usuarios.

Por lo tanto, este entorno considera al usuario como fuente de información y se nutre del nuevo papel principal que adquiere y de su nueva capacidad de decisión, ya que, como se ha señalado, el usuario no sólo lee, sino que ya crea, opina y difunde información de forma activa.

³ En adelante TICs

Pero no todo son ventajas en esta nueva “era digital”. Existe el problema de la sobrecarga informativa o “infoxicación”, donde el usuario debe enfrentarse a un volumen ingente de información que genera saturación y dificulta el proceso de búsqueda, recuperación y tratamiento de información. Como explicaba Cornella (1999) la “*infoxicación*” da lugar a una intoxicación debido al exceso de información imposible de asimilar.

Ante esto, recordemos que ahora es el propio usuario el que genera y aporta información a la web y en consecuencia es él el que también puede contribuir a resolver este problema gracias a la descripción, clasificación y organización adecuada de sus propios contenidos, mejorando la estructura de la información web. En conclusión, el etiquetado de contenidos nace para poder categorizar y organizar los contenidos creados en este nuevo entorno y ser una vía que facilite la recuperación de información optimizando tiempo y recursos, frente al crecimiento exponencial y desorganizado de la información digital.

Tradicionalmente han sido los profesionales de la información los encargados de describir y clasificar contenidos mediante vocabularios controlados o mediante una indización automática donde a través de algoritmos se detectan los términos más relevantes o con mayor peso del recurso. En cambio, ahora son los autores/usuarios los que comienzan a ocuparse de esta tarea, asumiendo el papel de “etiquetadores”. De esta forma, la indización profesional y la utilización de vocabularios controlados pasa a un segundo plano, y surge un etiquetado libre en lenguaje natural que carece de estructuras conceptuales o epistemológicas predefinidas que determinen su creación o asignación.

2.2.1 Etiquetado libre

Una etiqueta o *tag* es un término o conjunto de términos creados de forma espontánea y en lenguaje natural, que describen e identifican el recurso al que es asignado, con el objetivo de orientar tanto a su creador como a otro usuario a encontrar la información que desea, es decir, favorece la clasificación, localización y recuperación de información digital. Han surgido y tomado protagonismo a partir de la indización libre en entornos webs colaborativos y marcadores sociales. Las etiquetas surgen como nuevos modelos de representación de contenidos en la Web permitiendo una organización y gestión social de la información digital (Rodríguez Roche & Pérez Sanchidrián, 2014). Además, es un hecho que la utilización de palabras clave o etiquetas en la recuperación y organización de información digital es una de las claves de la Web 2.0.

El procedimiento en el que se asigna una etiqueta a un recurso para caracterizarlo o describirlo se denomina etiquetado o *tagging*. Esta modalidad adquiere la cualidad de libre, al utilizar términos que provienen del propio lenguaje o vocabulario cotidiano del usuario etiquetador no experto, es decir un lenguaje natural que no se somete a ningún control morfológico o semántico. El etiquetado libre o *tagging* se presenta como un nuevo tipo de indización en lenguaje natural, donde no es necesario tener ningún conocimiento teórico previo (Rodríguez Yunta, 2009). Frecuentemente, en diferentes recursos webs (blogs, páginas web, OPACs, etc.) y sistemas de recuperación de información se utilizan nubes de etiquetas o “*tags cloud*”, como recurso visual, para ofrecer las etiquetas al usuario que han sido asignadas con anterioridad y teniendo en cuenta la frecuencia de utilización de las mismas.



Imagen 2. Nube de etiquetas del catálogo del Sistema Nacional de Bibliotecas de Euskadi

Según Moens (2002 citado en Saavedra Saldívar, 2010), las principales ventajas de las etiquetas asignadas mediante un etiquetado libre son:

- Ofrecen una mayor expresividad y flexibilidad en la descripción, proporcionando varias perspectivas de un mismo recurso y aumentando sus puntos de acceso.
- Al no existir criterios o reglas para su creación o asignación, las etiquetas representan conceptos nuevos y complejos fácilmente, dando lugar a un lenguaje evolutivo e inmediato.
- Se presentan sin patrones relacionales o estructuras jerárquicas, ofreciendo una enorme riqueza de vocabulario y favoreciendo la recuperación de información.
- Las etiquetas son menos estrictas y se adaptan al recurso, al contrario que los descriptores fijos de un vocabulario controlado.

Golder & Huberman (2006) establecen una tipología de etiquetas en función de su uso y aplicación:

- Etiquetas “*Who Owns It Is*”: indican al autor o creador del recurso. Ejemplo: OMS, Unesco, IFLA...
- Etiquetas “*What or Who it is about.*”: identifican el tema del que trata el documento. Ejemplo: anatomía, tuberculosis, Lewis Carroll...
- Etiquetas “*What it Is*”: identifican qué es el recurso. Ejemplo: tesis, videoblog, referencia...
- Etiquetas “*Self Reference*”: dirigidas hacia el propio etiquetador (para su beneficio personal). Ejemplo: “mis comentarios” o “mis resultados”.
- Etiquetas “*Task Organizing*”: vinculadas con la realización de tareas. Ejemplo: “para leer” o analizar.

Quispe Solis (2013) interpreta que los usuarios asignan etiquetas basadas en su propio lenguaje, sin estar sometidas a jerarquías o estructuras preestablecidas, al contrario que los tesauros u otros vocabularios controlados. Además destaca de este tipo de indización su carácter evolutivo y actualizado ofreciendo una enorme riqueza de vocabulario.

2.2.2 Etiquetado social

El proceso de indización en lenguaje natural se puede realizar de forma individual o personal, es decir, realizado por un único usuario, o de forma social y colaborativa, en el que participa un grupo de ellos. Este último es el que adquiere un papel más importante en la actualidad, además de ser considerado como el impulsor del etiquetado de información en el contexto digital.

En el etiquetado social o *social tagging*, la descripción del recurso se obtiene por agregación, donde un conjunto de usuarios etiquetan un mismo recurso, obteniendo una descripción realizada desde diferentes puntos de vista y llevada a cabo en una dimensión colaborativa y democrática, donde el individualismo se diluye (Hassan Montero, 2006). Esta modalidad de descripción también denominada por McFedries (2006) como categorización popular (*Folk categorization*), etiquetado popular (*Folk tagging*) o etiquetado abierto (*Open tagging*), mejora la accesibilidad y localización de información proporcionando resultados más precisos y enriquecidos. Además, el fenómeno del etiquetado social aumenta cada vez más gracias a la adaptación de las necesidades del usuario.

Otra tipología vinculada con el etiquetado social es la que ofrece Cañadas (2006 citado en Yedid, 2013) y que surge en función de las motivaciones y objetivos de los usuarios al asignar etiquetas dentro de un entorno colaborativo:

- Etiquetado Egoísta: corresponde a la utilización de etiquetas muy significativas para el etiquetador, pero no para el resto de la comunidad, por lo que fuera del contexto personal las etiquetas no llegan a ser del todo relevantes. Como indican Pérez Sanchidrián, Campos Posada & Campos Posada (2014) la construcción de las etiquetas depende de la madurez del usuario. Finalmente, este tipo de etiquetado disminuye del beneficio social e incrementa el ruido informacional.

El etiquetado personal (indización realizada por una única persona) es de carácter egoísta, pues crea sus propio índice etiquetas para sus recursos, Hassan Montero (2006) denomina esta acción como "*personomía*".

- Etiquetado Amiguista: se etiqueta para compartir con personas de la misma comunidad o grupo (amistades, compañeros de clase, miembros de un departamento, etc.). El beneficio social es mayor que en el anterior tipo, aunque las etiquetas solo son conocidas por los miembros de dicho grupo y ajenas al resto.
- Etiquetado Altruista: en este caso, no hay un beneficio personal y directo para el usuario que etiqueta, pero sí un beneficio social que mejora la recuperación de información. El objetivo es compartir con el mayor número de personas, para ello se escogen etiquetas que sean comúnmente conocidas (Saavedra Saldívar, 2010).
- Etiquetado Populista: consiste en etiquetar con la finalidad de dar a conocer el contenido o como su propio nombre indica hacerlo más popular. Por lo que se utilizan etiquetas que sean destacadas y llamativas para la comunidad.

Por último, existen unos criterios para realizar el etiquetado social donde los usuarios pueden centrarse en distintos aspectos (Gómez Díaz, 2012):

- Según el contenido: donde el usuario etiqueta describiendo en función de la temática y del contenido del propio objeto. Por ejemplo, si un trabajo trata sobre libros electrónicos, se pueden utilizar etiquetas como como “*ebook*” o “libro electrónico”.
- Según el contexto: el usuario etiqueta aportando información sobre el entorno o lugar donde se ubica el objeto que va a describirse. Por ejemplo, si se va a etiquetar una foto de la fachada de las Escuelas Mayores, “Universidad de Salamanca” podría funcionar como etiqueta.
- Según las sensaciones o sentimientos del etiquetador: el usuario etiqueta reflejando lo que el objeto a describir le produce. El alto carácter subjetivo reduce la utilidad de las etiquetas. Es importante destacar como la indización tradicional en ningún caso representa lo que el recurso produce al etiquetador.

Los siguientes criterios, son puramente personales y de aplicación subjetiva. Su utilización no es aconsejada en comunidades o entornos colaborativos:

- Etiquetado organizativo: donde se describen recursos propios del usuario, es decir, el objetivo del etiquetado es puramente personal.
- Etiquetado según la procedencia: no se llega a describir el contenido del recurso, sino que la etiqueta orienta sobre la procedencia u origen del mismo.

Es importante comprender que no hay un estilo fijo de etiquetado pues en función de las necesidades y objetivos del usuario estos pueden ir variando. Según Pérez Sanchidrián et al. (2014) en el etiquetado social las ventajas de los usuarios crecen con la agregación colaborativa de etiquetas, y disminuyen con una descripción enfocada hacia los interés personales del etiquetador; además, desde el punto de vista de la recuperación de la información, el etiquetado social ofrece mayor exhaustividad y consistencia en la indización.

2.2.3 Folksonomías

El resultado final del etiquetado libre y social se conoce con el nombre de *folksonomía* (en inglés *folksonomy*). Hace referencia al conjunto de etiquetas asignadas en lenguaje natural para la descripción colaborativa de recursos, llevada a cabo por cualquier individuo en un entorno abierto, social y democrático. Esta nueva herramienta se presenta como una alternativa a los lenguajes documentales, ya que las etiquetas recopiladas en una folksonomía no serían aplicadas en una indización tradicional realizada por documentalistas.

Las folksonomías se basan en la colaboración y agregación libre, apartándose de cualquier tipo de control y estructuras jerarquizadas y organizadas; por lo tanto la clasificación del contenido no se realiza a través de una serie de categorías fijas y preestablecidas como tradicionalmente se ha hecho, sino todo lo contrario, las etiquetas se añaden y gestionan libremente por los propios usuarios, reflejando el poder popular (Díaz Piraquive et al., 2009).

Este concepto fue propuesto por Thomas Vander Wal. Proviene del término en inglés *folksonomy*, fruto de la palabra *Folk* (popular) y taxonomía, derivada a su vez de *Taxis* (clasificación) y *Nomos* (ordenar), de modo que etimológicamente hace referencia a la “clasificación popular”(Díaz Piraquive et al., 2009). Vander Wal (2007) establece dos principios básicos de toda folksonomía: la etiqueta, como representación adecuada del contenido, y el objetivo y finalidad del proceso de etiquetado. Ambos fundamentales para la desambiguación de las etiquetas asignadas y proporcionar una rica comprensión de la información etiquetada.

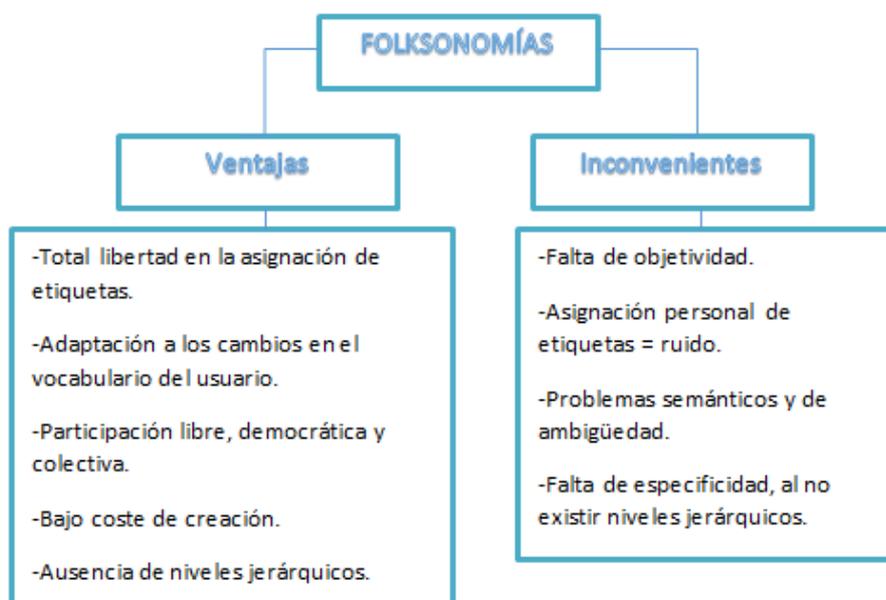


Imagen 3. Ventajas e inconvenientes de las folksonomías (Elaboración propia)

Finalmente, el propio Vander Wal (2005) propone una clasificación de las folksonomías, que ha conseguido una gran aceptación dentro de la bibliografía especializada, dividiéndolas en dos categorías:

- Genéricas o amplias (*broad folksonomies*): creadas de forma colaborativa, donde cada usuario etiqueta un mismo recurso a partir de su propio vocabulario e idioma. En esta tipología es común encontrar etiquetas populares debido a su alta visibilidad (teoría de la curva de poder o *power law curve*). Ejemplo: Delicious.
- Específicas o reducidas (*narrow folksonomies*): un número reducido de usuarios etiquetan el recurso, incluidos los creadores del mismo (*personomía*), utilizando etiquetas que los propios etiquetadores utilizarán para su posterior recuperación. Existe una menor riqueza terminológica y documental que en las genéricas. Ejemplo: Flickr.

2.2.3.1 Diferencias entre folksonomías y vocabularios controlados

Las folksonomías hacen frente al inmovilismo de los vocabularios tradicionales y controlados y su aplicación práctica ha sido rápida en toda la web. En el siguiente cuadro Soler Monreal & Gil Leiva (2010) exponen las principales diferencias entre la indización mediante folksonomías y la indización con vocabularios controlados, en este caso los tesauros:

Aspectos	Tesauros	Folksonomías
Objetivo	Representar y buscar información	Etiquetar información
Origen	Década de 1950	Década de 2000
Cobertura	Puede o no estar restringida a un campo del saber	General
Entorno	Analógico y digital	Digital
Entidades	Estables	Inestables
Fuentes	Autorizadas	Subjetivas
Productores	Expertos a partir de normas	Usuarios internautas
Auditoría/propiedad	Institución o persona que lo elabora	Inexistente
Lenguaje	Normalizado y controlado	Natural y dinámico
Estructura	Compleja	Simple
Coste de elaboración	Elevado	Bajo
Actualización	Periódica	Inmediata
Valor añadido	Organización conceptual de un campo del saber	Estudio del vocabulario usado por una comunidad y su evolución
Tipo de relaciones	Jerárquicas, asociativas y de equivalencia	Asociativas (co-aparición de palabras)
Usuarios	Profesionales de la información, requiere aprendizaje	Usuarios de internet, no requiere aprendizaje

Exhaustividad/precisión	Menor exhaustividad pero mayor precisión	Mayor exhaustividad pero menor precisión
--------------------------------	--	--

Tabla 1. Comparativa entre tesauros y folksonomías (Soler Monreal & Gil Leiva, 2010)

En conclusión, la popularidad y expansión de las folksonomías es indiscutible, y son cada vez más empleadas por todo tipo de usuarios y sistemas colaborativos, puesto que consiguen el propósito con el que surgieron: la organización de los contenidos y recursos web de una forma simple, rápida y barata, renovando la tarea de indizar y repartiendo la responsabilidad entre los usuarios/etiquetadores. Pendiente queda el estudio pormenorizado de su eficacia a la hora de la recuperación de la información, a términos de concreción y exhaustividad en los resultados de las búsquedas.

2.2.4 Lenguaje natural versus vocabularios controlados

Aunque el etiquetado libre sea una cuestión relativamente reciente, existen diferentes posiciones acerca de la descripción de recursos, pues este modelo se está planteando como un método popular o incluso como una herramienta básica para la organización de la información en internet. Y esto, simultáneamente crea un debate sobre si es viable permitir que los usuarios agreguen etiquetas sin que se empleen vocabularios que establezcan un nivel mínimo de organización y control, que garantice un adecuado tratamiento y posterior identificación de la información.

En contraposición a la descripción de información en lenguaje natural se encuentran Sánchez-Zamora, Llamas Nistal & Fernández Iglesias (2010) que destacan el problema de la ambigüedad (generado por el uso de sinónimos, homónimos, plurales y términos en distintos idiomas), como una de las principales inconvenientes del etiquetado libre, seguido de la falta de objetividad por parte del etiquetador. Para resolver esta cuestión recomiendan la posibilidad de establecer cierto control, pero entonces, la esencia del lenguaje natural disminuiría. Una idea parecida es la propuesta por Saavedra Saldívar (2010), que indica que mediante la alfabetización de los etiquetadores, es decir, estableciendo ciertas pautas generales, el etiquetador asignaría palabras clave evitando la utilización de términos que entorpezcan la descripción y recuperación de información.

Sin embargo, aunque exista la posibilidad de educar al usuario, las folksonomías siguen careciendo de estructuras y niveles jerárquicos, que orienten y guíen al usuario y permitan la expansión o restricción de consultas, al contrario que, por ejemplo, los tesauros. De modo que esta ausencia es un punto negativo frente a los vocabularios controlados (Heckner, Mühlbacher & Wolff, 2008).

Frente a esto, hay quienes argumentan que en las folksonomías no existen sinónimos, ya que los usuarios emplean las etiquetas por unas razones específicas, y que por lo tanto cada etiqueta utilizada tiene un significado único y personal. Aun así, es una realidad la presencia de sinónimos, términos ambiguos o faltas de ortografías que restan valor a las folksonomías puesto que provocan ruido e implican grandes limitaciones para la recuperación de información. Por lo tanto, se resalta la importancia de cierto control en este entorno libre, ya que mientras mayor sea la organización y estructuración, más fácil será la localización y recuperación de información (Díaz Piraquive et al., 2009).

Según Pérez Sanchidrián et al., (2014) dos aspectos indiscutibles de los esquemas tradicionales son la falta de actualización y adaptación a la forma de pensar del usuario, y su alto costo en comparación con las folksonomías, que se generan automáticamente y ofrecen etiquetas actuales y precisas capaces de representar cualquier tipo de contenido.

No obstante, profesionales de la Documentación han debatido sobre la utilización del lenguaje natural, y probablemente influidos por su estilo de trabajo, consideran que el control y la normalización son irrenunciables en la indización y que los vocabularios controlados deben seguir estando presentes, pues son los que eliminan ambigüedad y establecen relaciones conceptuales, al contrario que las folksonomías. Sin embargo, Moreira González (2007), destaca de los lenguajes natural, su carácter libre, colaborativo y democrático, aportando una enorme riqueza de vocabulario, y favoreciendo la recuperación de información. Además, expone que los vocabularios controlados y demás herramientas de clasificación deben renovarse para poder sobrevivir en este nuevo panorama.

De igual forma, Gómez Díaz (2012), apoya el etiquetado libre y social, y aleja la idea de ser una amenaza en el entorno profesional, pues supone un complemento muy enriquecedor, que favorece el acceso y mejora la visibilidad de la información. Además, habla sobre como la indización ha dejado de pertenecer en exclusiva a documentalistas y bibliotecarios, para ser responsabilidad de cualquier usuario de la red. Por lo tanto, desde su punto de vista, el etiquetado dentro del mundo de la Documentación, ha supuesto una ayuda considerable en todo el proceso de gestión de la información. Por último, defiende la necesidad de potenciar y aprovechar este nuevo papel del usuario, y seguir aprovechándose de todas las nuevas oportunidad que ofrece el aprovechamiento del conocimiento colectivo.

Finalmente, Soler Monreal & Gil Leiva (2010) plantean que los vocabularios controlados y el lenguaje natural no son incompatibles entre sí y sugiere la opción de utilizarlos conjuntamente, o sea que en la descripción de recursos podrían estar presentes varias formas de representación de información. Esta misma idea la sostenían Palhares Moreira & Murilo Stempliuć (2006), pues proponen desarrollar un nuevo modelo de indización, que tenga en cuenta la facilidad y simplicidad de la asignación de etiquetas en lenguaje natural, y la organización y estructuración de los vocabularios controlados, dicho de otro modo, pretenden recopilar los beneficios de ambos tipos, para asegurar un adecuado tratamiento de la información.

Dejando a un lado las ventajas e inconvenientes de cada tipo de vocabulario e indización, es importante asumir que el etiquetado libre en lenguaje natural es una realidad, y por lo tanto, sería interesante comprender la actitud de los usuarios ante este fenómeno y el porqué del uso de ciertas etiquetas, para ello más adelante se revisan los siguientes estudios realizados en esta misma materia.

2.3 La terminología en la Documentación

Como se puede observar durante el desarrollo de este trabajo, es innegable que la terminología tiene una influencia directa en las Ciencias de la Documentación, de la que se sirve y a la que sirve. Esta disciplina es clave para la representación, localización y acceso a la información. Además, la terminología se basa en la estandarización y normalización, factores esenciales para los profesionales de la información, ya que estos necesitan y trabajan con herramientas de unificación y control terminológico, para poder desarrollar tareas de indización y recuperación de

información. Estos son los lenguajes documentales, tesauros, ontologías y demás clasificaciones, formados por listados terminológicos, que describen un dominio específico y que están controlados formalmente (Cabré, 1995).

Los vocabularios controlados son mucho más que una simple recopilación de términos de una materia determinada. Gracias a la influencia de la terminología, los vocabularios controlados se fundamentan en la creación de relaciones conceptuales entre términos, para mejorar la organización, distribución y gestión del conocimiento mediante la aplicación de un único sistema terminológico (Moreno Ortiz, 2008). Por lo tanto, estas herramientas conceptuales controladas y normalizadas, facilitan la labor de gestión y tratamiento documental al bibliotecario.

Pero no solo son los bibliotecarios los que emplean vocabularios controlados para la recuperación de información; ahora motores de búsqueda, bases de datos y otros recursos web se nutren de estos lenguajes, estableciendo coincidencias entre la terminología utilizada en la consulta planteada por el usuario, frente a la terminología del sistema, para poder devolver resultados relevantes, e incluso ampliar estas consultas gracias a las relaciones terminológicas (asociativas, jerárquicas y equivalentes) establecidas en los vocabularios controlados. Además, estos instrumentos terminológicos son indispensables para eliminar los problemas lingüísticos propios de la descripción en lenguaje natural que generan ambigüedad y ruido documental y que aún no han sido resueltos. Boixados Sanuy (2012) destaca los siguientes:

- **Polisemia:** hace referencia a todos aquellos términos que tienen múltiples significados y que representan más de un concepto por razones contextuales. En la tarea descriptiva la polisemia entorpece y dificulta el proceso de recuperación de información obteniendo resultados no pertinentes.
- **Sinonimia:** son aquellos términos que comparten el mismo significado. Es uno de los problemas más destacados del etiquetado libre. La existencia de sinónimos en la descripción de contenidos genera un enorme ruido informacional que dificulta tareas como la búsqueda específicas de información. Este problema se incrementa en sistemas colaborativos o marcadores sociales, donde comunidades de usuarios etiquetan los mismos recursos sin ningún control terminológico.
- **Homonimia:** son aquellos términos que se escriben igual pero difieren en su significado. Este fenómeno está estrechamente vinculado con la polisemia y en sistemas colaborativos o marcadores sociales podrían resolverse con la utilización de términos relacionados ya preestablecidos en vocabularios controlados.
- **Categorización cognitiva o nivel de especificidad:** hace referencia al nivel de especificidad que se aplica en el momento de describir un recurso. En el caso del etiquetado en lenguaje natural el nivel de especificidad aplicado depende en cierta medida de dos cuestiones:
 - Del conocimiento que el usuario tenga respecto al contenido del recurso a describir: por ejemplo, investigadores describirán sus trabajos de una forma más específica y técnica, al tener un conocimiento avanzado de la materia.

- De los intereses del etiquetador: si el recurso es para compartirlo con otros usuarios, la descripción se realizará de una forma generalizada y utilizando términos comunes que eviten algún tipo de confusión en el grupo.

La ausencia de relaciones jerárquicas en la descripción en lenguaje natural, provoca cierta desorganización terminológica y conceptual. Al contrario que en los vocabularios controlados donde existen estructuras y niveles de especificidad, lo que da lugar a una correcta indización y organización del conocimiento.

- **Otras consideraciones:** existen otros problemas lingüísticos presentes en la terminología utilizada para la representación de información en lenguaje natural, como son la utilización indistintamente del plural o singular, mayúsculas y minúsculas, espacios en blanco o guiones para unir términos y errores ortográficos. Al igual que en el caso anterior, estas variaciones morfológicas se resuelven con la utilización de lenguajes documentales.

Todo esto influye directamente en la eficacia a la hora de localizar, gestionar y recuperar información, poniendo en riesgo la utilidad descriptiva de las palabras clave o etiquetas. En consecuencia, es necesario contar con el control terminológico existente en los vocabularios controlados, para que los anteriores problemas queden resueltos.

En conclusión, la terminología sirve a la Documentación para organizar y unificar el conocimiento (por medio de normas y estándares), lo que facilita su gestión gracias a la utilización de herramientas controladas terminológicamente (Cabré, 1995).

2.4 El uso de etiquetas en unidades de información

En la actualidad las etiquetas están presentes en distintos tipos de unidades de información y entornos webs. Como se expone anteriormente las palabras clave surgieron para la indización de información en bibliotecas, pero con los avances tecnológicos sus aplicaciones y funcionalidades han ido variando.

En un inicio las palabras clave se asignaban de forma manual por profesionales, pero más adelante este proceso se automatizó mediante sistemas que utilizaban algoritmos de frecuencia, relevancia, etc. Toda esta evolución ha ido acompañada de la utilización de vocabularios controlados para la indización tanto manual como automática, pero con la llegada de la Web esto varía, dando al usuario la capacidad de asignar sus propias etiquetas sin ningún control terminológico y sin limitaciones en el lenguaje. Esto ha creado una gran polémica entre profesionales de la información e investigadores de esta y otras materias.

Muchos catálogos y OPACs, añaden nube de etiquetas (creando folksonomías) en sus páginas principales para resaltar las búsquedas más realizadas por otros usuarios. También permiten que los usuarios describan y clasifiquen recursos, para poder compartir con otros, o simplemente para que a la hora de buscar información, el usuario se encuentre con etiquetas asignadas en un lenguaje más “conocido” y “cotidiano”, al contrario que si se hubieran utilizado vocabularios controlados, formados por términos más técnicos y especializados que quizá el usuario desconoce. Un ejemplo es el OPAC de la biblioteca digital del banco BBVA llamado Infobook. En la imagen 4, se muestra como en la interfaz de búsqueda existe la opción de buscar por

“tags”⁴, y accediendo encontramos una nube de etiquetas que redirigen al usuario a los recursos que tienen asignados la etiqueta seleccionada. Además, Infobook también permite la asignación de etiquetas por el usuario sin ninguna restricción ni limitación terminológica (Imagen 5).

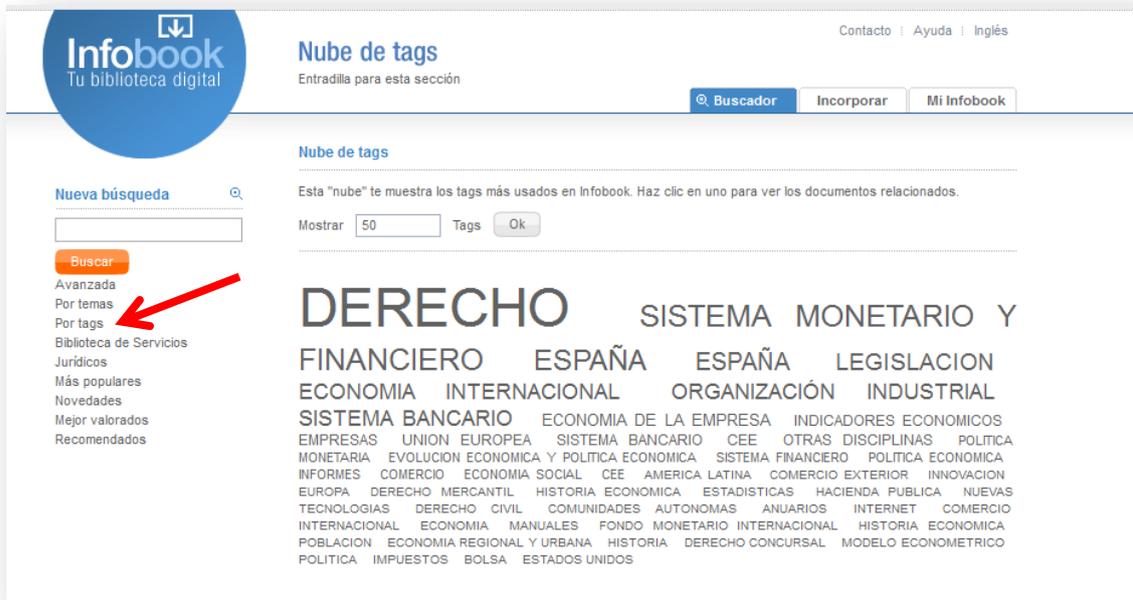


Imagen 4. Interfaz del OPAC de la biblioteca digital Infobook, destacando la opción de búsqueda “Por tags”



Imagen 5. Registro de un recurso en Infobook, destacando la sección de tags asignadas por los propios usuarios

⁴ “Tags” es el término en inglés de etiquetas (en singular tag)

Otro ejemplo es CIELO⁵ la plataforma de préstamo de libros electrónicos de la Universidad de Salamanca. Ofrece la funcionalidad de etiquetado de recursos por parte de los usuarios previamente registrados. En la siguiente imagen se observa un registro determinado donde usuarios libremente han asignado etiquetas. CIELO también permite la navegación a partir de las mismas, ya que al acceder a una de ellas se reenvía al usuario a todos los recursos descritos por dicha etiqueta, favoreciendo la navegación y ahorrando tiempo en la tarea de búsqueda de libros electrónicos.

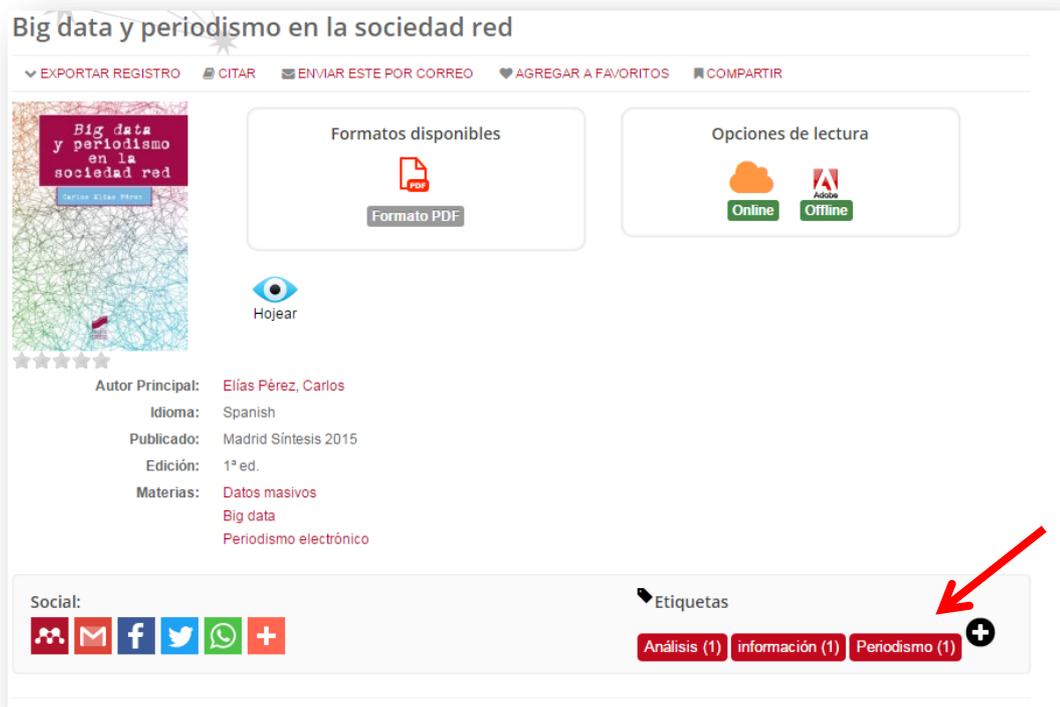


Imagen 6. Registro de un libro electrónico en CIELO, destacando las etiquetas asignadas por los propios usuarios

Conviene destacar que, aunque el usuario tiene total libertad a la hora de asignar etiquetas, CIELO recomienda la utilización de comillas en las etiquetas compuestas por varias palabras. Esta práctica de normalización favorece y mejora la organización y localización de recursos (Imagen 7).

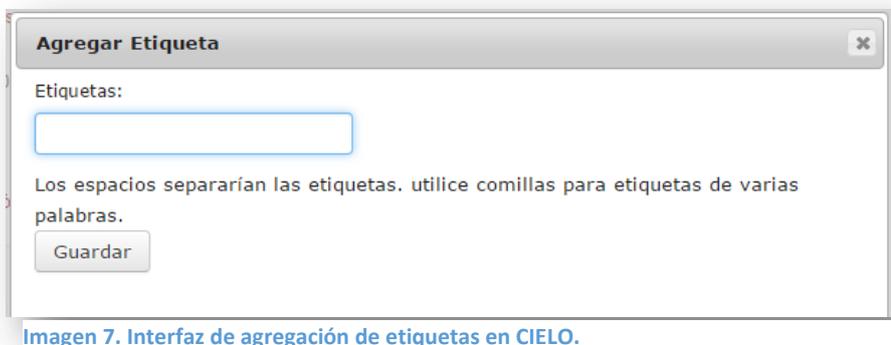


Imagen 7. Interfaz de agregación de etiquetas en CIELO.

⁵ Accesible en: <http://cielo.usal.es/>

El metabuscador bibliográfico WorldCat ⁶ también se beneficia de las funcionalidades del etiquetado libre y social. Una vez registrado el usuario puede asignar etiquetas a los registros bibliográficos existentes. En la imagen 8 se observa como esta base de datos permite mostrar las etiquetas en formato lista y en nube (útil cuando existe un número alto de etiquetas asignadas).

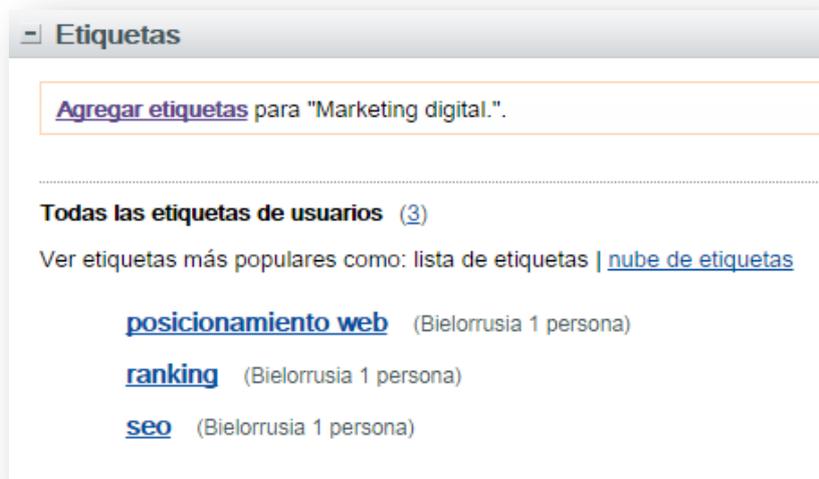


Imagen 8. Sección de etiquetado dentro de un registro bibliográfico en Worldcat

Además, el usuario dispone de un espacio personal donde puede gestionar (editar, eliminar, compartir) las etiquetas asignadas (Imagen 9).

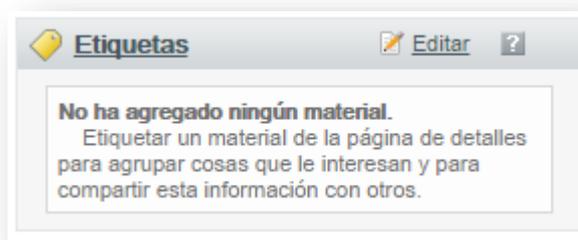


Imagen 9. Gestión personal de etiquetas en Worldcat

Por último, el catálogo de la biblioteca del distrito Ann Arbor ⁷ muestra las etiquetas de cada registro con un mayor o menor tamaño tipográfico, en función de su popularidad y uso (imagen 10). Este medio visual otorga unos valores y pesos a las etiquetas según su relevancia, lo que facilita la búsqueda a los usuarios.

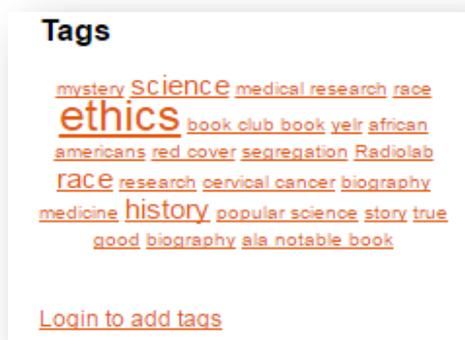


Imagen 10. Nube de etiquetas en un registro de

⁶ Accesible en: <http://www.worldcat.org/>

⁷ Accesible en: <http://www.aadl.org/>

Saavedra Saldívar (2010), defiende que el etiquetado ha aportado un valor añadido a las bibliotecas, hemerotecas y centros de documentación digitales, pues posibilitan un acercamiento directo entre el usuario y el bibliotecario gracias al *feedback*, y esto encamina hacia una mejora de los servicios bibliotecarios, agilizando y facilitando el proceso de análisis de contenidos, sin olvidar la mejora en la navegación y difusión de información.

Plataformas como *Delicious*, *SlashDot*, *Flirck*, *Technorati*, etc., han popularizado el etiquetado libre y social, como nuevas herramientas de clasificación y descripción de recursos. Estos sistemas gratuitos de gestión de información con conocidas como “bookmarking” o “marcadores sociales” y permiten al usuario gestionar recursos webs (urls, vídeos, imágenes, noticias, etc.), y añadirles etiquetas para su futura recuperación y para la utilización de estos recursos por otros usuarios del sistema. La utilización conjunta de sistemas colaborativos o marcadores sociales junto con la nueva tendencia del etiquetado libre da lugar a espacios con cierto enfoque semántico (Pérez Sanchidrián et al., 2014). Pero el uso de etiquetas no solo está presente en unidades de información relacionadas con el ocio o entretenimiento, existen sistemas destinados a la catalogación de recursos, como es *Librarything*, donde el usuario asigna etiquetas, con el objetivo de crear su biblioteca personal, o la herramienta *Mister Wong*, donde se etiquetan recursos educativos, para su difusión y divulgación.

En conclusión, el uso de etiquetas para la descripción de contenidos en unidades de información, presenta las siguientes particularidades:

- Mejora la organización en unidades de información.
- Facilita la búsqueda, navegación y recuperación interna.
- Proporciona información extra recomendada por otros usuarios/etiquetadores.
- Mejora la navegación y difusión de información.

Finalmente, hemos visto como el etiquetado está presente en cada vez más unidades de información. Pero aún podría ocupar una posición más relevante dentro de empresas y organizaciones privadas. Las folksonomías en empresas funcionan como nexo entre usuarios/empleados y ofrecen un valor añadido mejorando la distribución de servicios, tareas de gestión de información y la comunicación interna. La fácil y barata creación y utilización de estas herramientas, surgen como alternativa de las tradicionales taxonomías institucionales que no se adaptan a los cambios del vocabulario del empleado. Por lo tanto, ya no es imprescindible que las empresas y organizaciones gasten grandes sumas de dinero en programas de gestión de información: con la incorporación de folksonomías a su ámbito de trabajo se reducirá el tiempo de dedicación de ciertas tareas y mejorará la conexión y colaboración interna dentro de grupos de trabajo (López García & Aneiros Díaz, 2009).

2.5 Descripción de recursos científicos

Hemos visto como la descripción de información mediante palabras clave comenzó con la indización de fondos bibliotecarios, pero con la llegada de las TICs el etiquetado se ha ido adaptando a nuevos recursos y formatos que han ido surgiendo. Empezó siendo más común aplicar etiquetas en blogs, videos, imágenes, etc., pero la comunidad científica y académica también ha comenzado a aprovecharse de las utilidades y beneficios del etiquetado por parte del autor/usuario.

Un ejemplo es Mendeley, un gestor de referencias bibliográficas gratuito, donde el usuario puede crear una base de datos personal pues permite almacenar y organizar trabajos. Además, tiene un propio buscador científico, creado gracias a la comunidad científica que integra sus propios estudios. Como gestor de referencias, Mendeley permite al usuario añadir etiquetas en lenguaje natural que sean significativas para él, además de mostrar las palabras clave asignadas por el autor de dicho trabajo (Imagen 6).

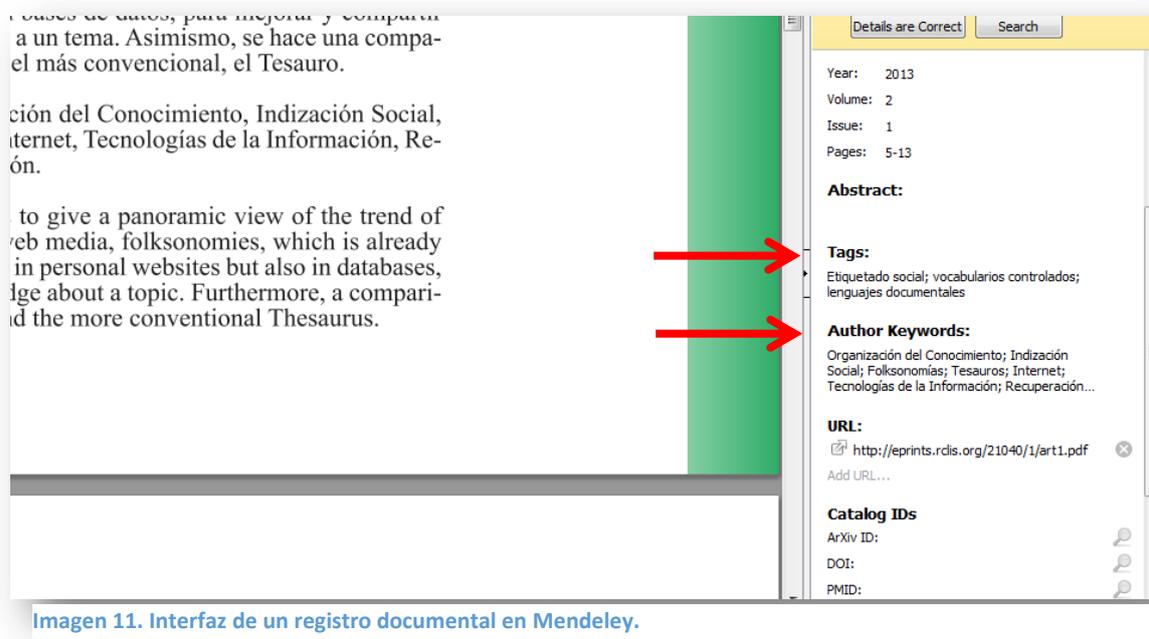


Imagen 11. Interfaz de un registro documental en Mendeley.

Connotea y Citeulike, son otras plataformas similares a Mendeley, muy utilizadas por la comunidad científica. De ambas herramientas se destaca la funcionalidad “Related tags”, que recomienda etiquetas relacionadas o parecidas a las añadidas por el usuario y la aplicación “Related users”, que agrupa usuarios que asignan las mismas etiquetas, lo que puede significar que estos compartan los mismos intereses y temas de estudio (Alonso Arévalo, Cordón García, & Martín Rodero, 2010). Esta última funcionalidad está estrechamente vinculada con los sistemas de recomendación, ambos crean grupos de usuarios que tienen un perfil similar, filtrando y rastreando sus intereses para poder ofrecerle un conjunto de recomendaciones personalizadas y adaptadas a sus necesidades. De esta forma, se consigue un servicio totalmente adaptado al usuario, pues este ya no tiene que buscar información, si no que con funcionalidades como “Related users” o los sistemas de recomendación, la información llega a ellos de una forma rápida y fácil y enriqueciendo la experiencia del usuario en estos entornos.

Existen otras plataformas de literatura académica y científica como son BibSonomy, Medline, Academia, etc., que permiten la descripción de contenidos. La clave del etiquetado en estas herramientas está en que la descripción de recursos científicos se hace en un inicio para el beneficio del usuario etiquetador, y los beneficios hacia el resto de usuarios son el resultado de este comportamiento. Gracias a este etiquetado individual se generan “Índices de popularidad”, un conjunto de etiquetas agrupadas por categorías científicas, que son útiles para hallar tendencias en diferentes dominios científicos. La idea es que otros profesionales observen las etiquetas más utilizadas en su disciplina, y puedan encontrar trabajos similares o incluso acceder a la colección de otros compañeros, y descubrir recursos útiles para sus investigaciones, de una forma más rápida y sencilla, ya que en la mayoría de los casos la popularidad está directamente vinculada con la calidad. Por lo tanto, la descripción de recursos científicos facilita el descubrimiento e incrementa la visibilidad de la producción científica (Alonso Arévalo, Cordón García, & Martín Rodero, 2012).

Otro ejemplo dentro del ámbito de la investigación, son las bases de datos científicas. Web of Science es una plataforma web que recoge bases de datos relevantes en diversas disciplinas. Esta indiza las palabras clave asignadas por los autores de los trabajos ya incluidas en las revistas, mostrándolas en los registros de los mismos. También, permite la búsqueda de estos términos a partir de un campo específico para ello (normalmente llamado *Keyword*). Además, gracias a la posibilidad de buscar en lenguaje natural, las bases de datos generan perfiles individuales junto con listados de palabras clave, para ofrecer sugerencias y recomendaciones, mejorando la experiencia de búsqueda del usuario gracias a unos resultados más personalizados.

Alonso Arévalo (2009), expone que al asignar y compartir etiquetas, el descubrimiento científico entre áreas de conocimiento es mucho más fácil. Además, gracias a la especificidad del etiquetado dentro de áreas especializadas, se reduce la posibilidad de que las etiquetas tengan varias interpretaciones, lo que otorga un mayor valor intrínseco al etiquetado. En conclusión, la descripción de recursos científicos ofrece un valor añadido en las tareas de investigación.

2.6 El etiquetado social y su aplicación documental: estudios previos

Dejando a un lado, las ventajas e inconvenientes de vocabularios controlados y folksonomías, es importante entender que el etiquetado en lenguaje natural es una realidad, y por lo tanto, sería interesante comprender la actitud de los usuarios ante este fenómeno y el porqué del uso de ciertas etiquetas o tendencias. Para ello, a continuación se revisan diversos estudios previos realizados sobre esta materia.

Heckner et al. (2008), han desarrollado tres modelos para analizar el comportamiento del usuario/etiquetador en Connotea, un gestor de referencias en línea para científicos e investigadores, donde los usuarios etiquetan artículos para su posterior localización y uso. Estos modelos se diferencian en tres categorías: Categoría lingüística (LTCM), Categoría Funcional (FTCM) y Categoría “Tag to Text” (T2TCM), aplicadas a 500 artículos seleccionados al azar. Entre la variedad de resultados obtenidos, se destaca que los usuarios utilizan y reciclan etiquetas ya usadas previamente. Este comportamiento “conservador” del usuario, está influenciado posiblemente por la forma en que el sistema y su interfaz muestran las etiquetas ya asignadas. Además, estos autores citan en su estudio a Sen et al. (2006), el cual indica que el usuario puede etiquetar bajo una tendencia personal, a partir de

su experiencia y conocimientos previos; o bajo la influencia de la comunidad, donde los usuarios actúan de igual forma que los demás miembros. Teniendo en cuenta esta tendencia personal, en el estudio también se detectaron etiquetas usadas con bastante frecuencia por solo un usuario, a las que se denominaron “exclusivas”, pues solo son significativas para el propio etiquetador. Otro resultado llamativo, es la creación de una estructura de clasificación personal, organizando así la información de una manera jerárquica, es decir, que dentro de la libertad existente, el usuario organiza y estructura sus etiquetas. Por último, este estudio realiza una distinción entre las palabras clave de autor del recurso y las etiquetas que asignan otros usuarios al mismo, observando como los autores intentan ser lo más específicos posible al describir el contenido de su trabajo y utilizan etiquetas más técnicas, en comparación con el usuario.

Previamente, Mathes (2004) ya observó cómo los usuarios reasignan y reciclan etiquetas, afirmando que la visibilidad de las mismas influye en su uso, puesto que las más utilizadas son las que están visibles y disponibles en el sistema (debido a su popularidad), y por lo tanto son reutilizadas por otros usuarios. Es evidente que la organización y diseño de los sistemas y herramientas de marcado, intervienen en el comportamiento del usuario en el momento del etiquetado. Asimismo, existe una tendencia hacia el “mínimo esfuerzo” por parte del usuario, al no crear nuevas etiquetas y utilizar las ya existentes. Este último concepto, también observado en el trabajo de Chi-Shiou & Yin Fan (2012) da lugar a la creación de un nuevo tipo de etiquetado, llamado “Etiquetado de Imitación”, definido como la acción de asignar etiquetas ya creadas con anterioridad.

Sin embargo, dos años más tarde Guy & Tonkin (2006), observaron el uso de etiquetas en Delicious y Flickr, pero enfocando la reutilización de las etiquetas de una forma diferente a los anteriores investigadores. Los resultados indican que las etiquetas de un solo uso son mínimas, ya que componen sólo el 15% de la muestra, por lo que los usuarios prefieren reutilizar etiquetas debido a su popularidad, y no por falta de esfuerzo, es decir, prefieren etiquetas que ya han sido aceptadas previamente, y así, el usuario evita errores en la descripción propia del contenido. Del mismo modo ha asimilado Doerfel et al. (2016) sus resultados al examinar el marcador social Bibsonomy, concluyendo que esta es la esencia del etiquetado libre, beneficiarse de las agregaciones de los demás, lo que justifica la naturaleza social y colaborativa del etiquetado libre. Si es cierto que Guy & Tonkin (2006) exponen un inconveniente en la reutilización de etiquetas, ya que si desde un inicio la etiqueta popular es incorrecta, estos sistemas pueden llenarse de etiquetas ambiguas, etiquetas personales (resultado de un etiquetado egoísta), etiquetas con faltas de ortografía y con diferente número y género. Por esta razón, afirman que el etiquetado no es un reemplazo de los vocabularios controlados, pero destacan la gran utilidad de las folksonomías.

Recientemente, Rodríguez Roche & Pérez Sanchidrián (2014) han analizado terminológicamente las etiquetas de post de blogs sobre Cuba. A pesar de que en los resultados se indica que existe una media de entre 4 y 6 etiquetas por post, y que estas hacen referencia a la historia y cultura de Cuba, se resalta una apreciación que detallada en el estudio. Los autores observan cómo una misma etiqueta puede ser visualizada desde diferentes enfoques y puntos de vista, en función de la interpretación de cada usuario (poniendo como ejemplo la etiqueta “*religión*”). Esta cuestión hace que reflexionen al respecto, y afirmen que la interpretación y asimilación de las etiquetas dependerá en gran medida de la opinión individual de cada usuario.

Dejando a un lado el comportamiento del usuario ante el etiquetado, se encuentra el dilema de la utilización de vocabularios controlados o el lenguaje natural, en la representación de la información, y por consiguiente en su futura recuperación.

Bogers & Petras (2015) comparan el uso de etiquetas y vocabularios controlados, en la búsqueda de libros, utilizando las plataformas LibraryThing, Amazon, la Biblioteca del Congreso y Biblioteca Británica, como proveedores de datos. Para ello, se han recopilado solicitudes reales de información. El objetivo era estudiar cuáles de los dos tipos de vocabularios (libre y controlado) mejora el proceso de recuperación y búsqueda de libros. El hallazgo principal muestra que las etiquetas y los vocabularios controlados trabajan de forma similar ante la búsqueda de libros, existiendo solo una ligera ventaja de las etiquetas en lenguaje natural, posiblemente al proporcionar una descripción más específica del contenido de los libros. Además, intentaron agrupar ambos vocabularios para ver como trabajaban juntos y se comprobó cómo se complementan entre sí. Otros estudios como Soler Monreal & Gil Leiva (2010) y Palhares Moreira & Murilo Stempliac (2006), también afirman que la unión de ambas técnicas, pueden ofrecer mejores resultados, y el rendimiento en el proceso de recuperación de información aumentaría significativamente. Esto indica que las etiquetas y los términos controlados podrían coexistir dentro de todo el proceso documental.

Sin embargo, esta idea ha sido cuestionada anteriormente por varios estudios, al relacionar etiquetas con descriptores. Por ejemplo, Danielle & Schleyer (2012), comparan las etiquetas de CiteULike, con los encabezamientos de materia médicos (MeSH) indizados en MEDLINE, y observaron cómo ambos son bastante distintos semánticamente, pues pueden reflejar diferentes puntos de vistas. De igual forma sucede en el estudio de Tomás-Casterá, Sanz-Valero, Wanden-Berghe, & Landaeta (2009), donde al estudiar las palabras clave utilizadas en los artículos de las revistas especializadas en nutrición, indizadas en Red SciELO, y los Descriptores de Ciencias de la Salud (DeCS), se obtiene solo un 31,9% de coincidencia exacta entre palabras clave y descriptores. Otro resultado similar, es el 30% resultante de la comparación de las palabras clave de autor y descriptores, asignados a trabajos ubicados en las bases de datos ISOC, IME E ICYT (Gil-Leiva & Alonso-Arroyo, 2005). Por consiguiente, estos autores consideran que la indización controlada no debe ser sustituida por una indización libre, pues se ha observado, como existe una dificultad al recuperar información especializada, mediante la utilización de términos en lenguaje natural. En consecuencia, se antepone la utilización de vocabularios controlados, para garantizar un acceso eficiente de información, y se reafirma que el etiquetado no podría converger con vocabularios controlados, y mucho menos sustituirlo en la tarea de indización.

Este concepto también ha creado polémica entre profesionales de la Documentación, existiendo diferentes opiniones respecto al etiquetado libre como herramienta de indización. En el trabajo de Rodríguez Yunta (2009) se analiza el etiquetado en blogs especializados y marcadores sociales, realizado por documentalistas y bibliotecarios. Se recopilan un listado de etiquetas, posteriormente comparadas con los descriptores del Tesoro de Biblioteconomía y Documentación editado por el CINDOC. Finalmente, tras debatir entre las diferentes ventajas e inconvenientes de cada uno de los vocabularios, el autor observa en sus resultados como existe un uso excesivo de singulares y plurales, sinónimos que dificultan el acceso de otros conceptos más relevantes o la presencia de errores ortográficos. Aun así, también se han encontrado nuevos términos que no estaban presentes en el tesoro, como es el caso de los blogs analizados, donde el 43% de las etiquetas podrían incorporarse al tesoro. Lo mismo sucede en el caso del marcador social, pues sólo existe un 8% de coincidencia de las etiquetas analizadas, con los

descriptores del tesoro. Por lo tanto, como reflexión final, el autor propone un estrechamiento entre los vocabularios controlados y las folksonomías, para beneficiarse por ejemplo, del carácter evolutivo y actualizado del lenguaje natural, y ampliar la cobertura terminológica en los lenguajes controlados, y así, aumentar sus utilidades entre otras cuestiones. De forma similar, el etiquetado libre debería adquirir cierto control y establecer recomendaciones mínimas, para una correcta asignación e incorporación de nuevos términos y variantes léxicas. Un ejemplo de recomendaciones podrían ser las propuestas por Mejias (2005). Estas son:

- Asignar etiquetas teniendo en cuenta su posterior uso por el resto de la comunidad, es decir, con una finalidad social y evitando el etiquetado egoísta o la “*personomía*”.
- Utilizar plurales en la asignación de etiquetas. Evitará tener que comprobar el número gramatical de las etiquetas. No obstante, en algunas ocasiones, será necesario especificar el número, ya que puede variar su significado. Por ejemplo, el término “humanidad” como cualidad y “humanidades” como disciplina.
- Evitar las mayúsculas, excepto cuando sea necesaria para diferenciar etiquetas. Por ejemplo, sería correcto poner en mayúscula la etiqueta “Asturiana” como marca comercial, para diferenciarla de “asturiana” como gentilicio.
- Asignar etiquetas tanto específicas como generales, para describir un recurso. Es decir, que describan al recurso en términos muy específicos, pero también aplicando etiquetas en términos generales. Esta cuestión facilita la recuperación de información, ya que aunque algunas etiquetas podrían ser demasiado amplias para el beneficio personal, ayudarían a otros a usuarios a encontrar el recurso.
- Evitar guiones y puntos para la agrupación de términos. Estos conectores crean ruido y duplicidad de etiquetas.
- Si con anterioridad existen pautas de etiquetado establecidas por otros usuarios, es recomendable adoptarlas (siempre que tengan sentido común). Esto dará lugar, a un conjunto de buenas prácticas.
- Contribuir al mantenimiento de las etiquetas y por consiguiente de las folksonomías. Es recomendable repasar y actualizar cada cierto tiempo estos términos, lo que aumentará el valor de la colección.

En definitiva, aunque existan diferentes opiniones y posturas sobre el tema, la opción de fusionar ambos lenguajes es una cuestión a considerar si se quiere que tesauros, ontologías y otros vocabularios controlados, puedan representar un papel protagonista en el futuro, para llevar a cabo tareas como la descripción, organización y recuperación de información digital.

2.7 Repositorio institucional GREDOS

A continuación se ofrece una breve descripción de la plataforma académica que acoge los recursos científicos y académicos utilizados para el estudio de caso realizado. Estos son los TFMs alojados en el repositorio institucional GREDOS⁸.

Según la RAE, el concepto repositorio proviene del término en latín *repositorium* y hace referencia a un armario o alacena, es decir es un lugar donde de guarda algo.

Como ya se ha indicado anteriormente, la herramienta o servicio objeto de análisis de este trabajo es GREDOS. GREDOS es el acrónimo de Gestión del Repositorio Documental de la Universidad de Salamanca. Desde sus inicios en el año 2009, su función ha sido y es reunir, gestionar, preservar y almacenar todos los recursos en formato digital producidos o alojados por la USAL, para su posterior recuperación. Como se indica anteriormente GREDOS, es una plataforma que acoge toda la producción digital de la USAL, por lo que tiene carácter integrador, puesto que todo recurso creado en la institución puede estar disponible en el repositorio (si cumple unos requisitos previos)(Ferrerías Fernández & Casaus Peña, 2009).

Dspace es el software de código abierto utilizado para gestionar el repositorio, ya que nace con la finalidad de alojar y administrar colecciones digitales.

GREDOS estructura jerárquicamente sus contenidos en comunidades, subcomunidades, colecciones y documentos (con sus respectivos registros y descripciones). Permite la consulta en línea de recursos históricos, científicos, didácticos e institucionales, distribuidos entre sus cuatro comunidades principales (Gredos, 2009)⁹:

- Archivo institucional (23.281 registros): contiene documentación interna de la USAL y de carácter institucional y administrativo (documentos administrativos, fondos históricos, guías académicas...).
- Biblioteca digital (63.372 registros): recopila colecciones digitalizadas de documentos históricos y patrimoniales (manuscritos, libros antiguos, prensa histórica...).
- Repositorio científico (24.415 registros): contiene producción de los investigadores y departamentos de la USAL (tesis doctorales, artículos, congresos científicos...).
- Repositorio docente (3.148 registros): recursos de carácter didáctico generados en la universidad (trabajados de fin de grado y máster, materiales didácticos, tutoriales de bibliotecas...).

⁸ Accesible en: <https://gredos.usal.es/jspui>

⁹ Censo realizado el 23/04/2017



Imagen 12. Página inicial de GREDOS

La USAL promueve y apoya que la producción científica generada gracias a subvenciones públicas sea de libre acceso. De modo que gracias al compromiso de la universidad con la tecnología Open Access, todos los contenidos de GREDOS están disponibles en acceso abierto, es decir, permite una consulta libre y gratuita a texto completo con el fin de asegurar su máxima difusión, eliminando barreras políticas, técnicas o económicas. Igualmente, en el repositorio se preservan los derechos de autor de los objetos almacenados gracias a la licencia “Creative Commons”, por lo que el uso o reproducción de cualquier recurso ubicado en el repositorio documental deberá ser citado y reconocido adecuadamente (Ferrerías Fernández & Casaus Peña, 2009). Cabe destacar que gracias a la tecnología interoperable que utiliza, GREDOS aumenta su visibilidad cediendo sus metadatos para que sean recolectados y agregados por otros repositorios, bibliotecas digitales o motores de búsqueda especializados, como son OAlster, Hispana, BASE, Europea, Recolecta, etc.

Además de permitir el acceso completo y libre a todos los contenidos de la USAL, GREDOS ofrece otros servicios y funcionalidades:

- Servicio de alertas a partir de la funcionalidad “Mi GREDOS” para poder recibir un aviso cada vez que haya un documento nuevo dentro de la colección deseada.
- Estadísticas sobre uso, visitas y descargar de recursos ubicados en el repositorio.
- Difusión del contenido gracias a la exportación de registros a gestores de referencias, correo electrónico y redes sociales.
- Servicio de redes sociales (Facebook y Twitter) y blog Diarium¹⁰, para poder interactuar con el usuario e informar de noticias actualizadas.

¹⁰ Accesible en: <http://diarium.usal.es/GREDOS/>

- Aplicación Autoarchivo: servicio donde el usuario/autor puede depositar su trabajo en GREDOS. Después de la introducción de metadatos y aceptación por los responsables del repositorio.

Finalmente, en el “Ranking Web de Repositorios del Mundo” (edición de Febrero de 2017) donde se mide la visibilidad e impacto de repositorios, GREDOS alcanza la posición 140 entre los 2284 repositorios del ranking mundial. Ocupa también la posición 61 entre los 900 repositorios recogidos en el ranking europeo, y en el ranking de España se ubica en la 9ª posición.

Prácticas académicas en GREDOS

Durante el periodo del 27 de abril al 8 de junio del año 2016 (125 horas), la autora de este trabajo realizó sus prácticas académicas correspondientes al Máster en Sistemas de Información Digital en el repositorio GREDOS. Las principales tareas realizadas fueron:

- Incorporación de tesis doctorales mediante la aplicación de esquemas de metadatos Dublin Core.
- Búsqueda y recogida de información para la elaboración de metadatos, puesto que algunos registros estaban incompletos o le faltaba algún dato. Por ejemplo, en ocasiones las tesis doctorales no tenían asignadas las palabras clave que el autor debía añadir, por lo que desde GREDOS se asignaban las convenientes provenientes de vocabularios controlados.
- Normalización y modificación de autoridades.

Además de los alumnos en prácticas y el técnico encargado del repositorio, los bibliotecarios de cada facultad incorporan los trabajos correspondientes a las titulaciones impartidas en dicha facultad, respetando las palabras clave proporcionadas por el autor. Como se ha indicado en la justificación de este trabajo, durante la realización del mismo se observó cómo GREDOS no respetada en algunos casos las palabras clave de autor, y una de las razones es la necesidad de normalización asignando términos admitidos o descriptores que signifiquen los mismo que las etiquetas pero que describan mejor el contenido del TFM.

3 Metodología

Para el desarrollo de este trabajo se ha seguido un conjunto estructurado de etapas que han ido desde la definición de los objetivos, pasando por una revisión bibliográfica, hasta la elaboración de una metodología clara para el desarrollo de un estudio de caso. Por lo tanto, en esta investigación se pueden diferenciar dos partes principales:

1. Exploración y revisión de la bibliografía del tema a tratar.
2. Estudio de caso, donde se lleva a cabo un análisis de las palabras clave de autor (recopiladas en el Anexo I), presentes en TFM's ubicados en el repositorio GREDOS.

3.1 Revisión bibliográfica

Como punto de partida, se ha realizado una revisión bibliográfica fundamental para el desarrollo del estado de la cuestión. Se ha comenzado con una bibliografía recomendada y proporcionada por el tutor, y otra, resultado de búsquedas personales.

En relación con las búsquedas personales, se han explorado diferentes bases de datos científicas:

- EBSCO: bases de datos que ofrece recursos a texto completo, índices y estadísticas. Abarca diferentes áreas de conocimiento, como medicina, económica, informática, educación, física, etc.
- Dialnet: es un portal bibliográfico de acceso libre creado por la Universidad de la Rioja. Hace visible la producción científica hispana, centrándose en el ámbito las Ciencias Humanas, Jurídicas y Sociales.
- Google Scholar: es una herramienta de Google que permite la búsqueda a texto completo de bibliografía especializada de un gran número de disciplinas y fuentes relevantes dentro del mundo de la investigación académica.
- Web of Science (WOS): es un servicio en línea de información científica, que permite el acceso a un conjunto de bases de datos y facilita la búsqueda de información multidisciplinaria de investigaciones actuales y retrospectivas.
- GREDOS: plataforma que acoge toda la producción científica y digital de la Universidad de Salamanca, y permite la consulta de recursos digitales de carácter histórico, científico, didácticos e institucionales.

El proceso de búsqueda se caracterizó por la combinación de términos y conceptos planteados tanto en castellano como en inglés: Etiquetado libre (*free tagging*), Indización libre (*free indexing*), Etiquetado colaborativo (*collaborative tagging*), Web social (social web), Palabra clave (*keyword*), Comportamiento usuario (*user behaviour*), Folksonomías (*folksonomies*), Vocabularios controlados (*controlled vocabulary*) y Sistemas de etiquetado (*tagging systems*).

Localizados los recursos relevantes para el estudio, se procedió a su análisis para esclarecer la evolución, funcionalidad y características que presenta la investigación de este tema.

Finalmente, en lo que se refiere a la bibliografía, cabe destacar que se ha utilizado el estilo APA Sexta Edición, tanto para el caso del material impreso como para los recursos digitales.

3.2 Estudio de caso

Tras haber finalizado el marco teórico de este proyecto, se comenzó con el estudio empírico. Al querer vincular este trabajo con el repositorio GREDOS, el objetivo inicial era encontrar una colección donde el repositorio mostrara las palabras clave asignadas por autor.

Al ser conocedores de que en los TFMs del presente máster es obligatoria la asignación de palabras clave por el autor, se utilizó esta colección como punto de partida. En el acuerdo de la Comisión Académica del Máster en Sistemas de Información Digital donde se determinan las cuestiones relativas a la presentación de TFMs, se informa de lo siguiente: *“El Reglamento sobre Trabajos Fin de Máster de la Universidad de Salamanca, dispone en su artículo 9.1 que cada Comisión Académica determinará anualmente las normas de estilo, extensión y estructura del TFM”*. Esto quiere decir que cada máster dispondrá de unas normas de presentación determinadas y sujetas a posibles cambios, es decir, puede o no existir la obligación de añadir palabras clave a los trabajos. Visto lo dispuesto en el reglamento se observó como en las indicaciones aprobadas por la comisión de este máster se establece que el alumno/autor debe entregar una ficha catalográfica adaptada al formato del repositorio institucional GREDOS, donde se deben incluir un conjunto de palabras clave en español y en inglés. Partiendo de esta información se realizó un breve muestreo inicial en la colección de TFMs de GREDOS, para observar si en otros másters se asignaban palabras clave. Al obtener un resultado favorable se decidió utilizar la colección de TFMs para el análisis de palabras clave de autor.

Además, en el muestro inicial también se comprobó como en algunos casos GREDOS no respeta las palabras clave de autor asignadas en los TFMs, y utiliza términos admitidos que significaban lo mismo. Esta cuestión incremento el interés de análisis en la colección de TFMs, ya que nos pareció útil estudiar la coincidencia de las palabras clave de autor con las que aparecen en el registro del repositorio.

Se ha utilizado el programa de hojas de cálculo Microsoft Excel 2010, para la recopilación y análisis de las etiquetas y para llevar a cabo el procesamiento de los datos métricos, así como la elaboración de tablas y gráficos. El conjunto de etiquetas analizadas han sido recopiladas en el Anexo I.

Una vez determinada la población que va a ser analizada, el paso siguiente fue determinar cuáles eran los aspectos que debería se deberían estudiar para poder obtener unos resultados lo más completos posible. Para la delimitación de las características y variables a analizar, se tomó como modelo y punto de partida el trabajo de Markus Heckner, Susanne Mühlbacher & Christian Wolff (2008). Además de utilizar algunos de los patrones de análisis que establece el anterior trabajo, se

añadieron otras características de análisis, en función de las necesidades que iban surgiendo durante el estudio de las primeras palabras clave recopiladas.

3.2.1 Variables de estudio

Con el objetivo de poder sistematizar las palabras clave de una forma eficiente para su posterior análisis, se establecieron seis grupos de características a analizar. A continuación se presentan los grupos y variables, junto con los criterios metodológicos establecidos para cada una de ellas:

1. **ETIQUETAS SEGÚN EL NÚMERO DE TÉRMINOS QUE LA FORMAN:**

1.1. **Etiqueta simple:** Etiqueta formada por un solo término o vocablo.

- Tipos de etiquetas simples:
 - Nombre común
 - Adjetivo
 - Adverbio
 - Número: tomarán valor las etiquetas formadas por dígitos (ej. 2015), y las escritas (ej. dos mil quince).
 - Verbo
 - Acrónimo: los acrónimos formados por dos términos seguirán considerándose etiqueta simple. Ejemplo: *STATIS DUAL* o *ISBD (ER)*.
 - Nombre Propio: todo nombre propio se ha considerado etiqueta simple aunque esté formado por más de un término, al estar referido a un solo concepto. Entidades, instituciones u organizaciones se han considerado nombre propio.

1.2. **Etiqueta compuesta:** Etiqueta formada por dos o más términos/vocablos, independientemente del sistema de separación que utilicen (guion, barra, etc.). Ejemplos: *Empresario-inversor*, *rentabilidad-riesgo*.

- Cuando una etiqueta compuesta, como por ejemplo, "*Tecnologías de la información y comunicación*" haga uso de su acrónimo, es decir "*TICs*", durante el trabajo, dicho acrónimo no será considerado para el análisis, solo se tendrá en cuenta la etiqueta compuesta.
- Cuando etiquetas formadas por un acrónimo, seguido del desarrollo de sus siglas (ejemplo: *(TIC) "Tecnologías de la información y comunicación"*), estas no serán consideradas como etiquetas compuestas, si no que tomarán valor en la variable "acrónimo".
- Las preposiciones y otros conectores (artículos, determinantes, etc.) no se han tenido cuenta como etiqueta.

2. CLASIFICACIÓN LINGÜÍSTICA

2.1. Etiqueta en otro idioma diferente al castellano

- Los acrónimos formados por palabras en otro idioma (diferente al castellano), toman valor en esta variable.
- En ocasiones los TFMs presentan el resumen y las palabras clave en otro idioma diferente al castellano, en estos casos solo se han analizado las etiquetas en castellano.
- Los nombres propios en inglés también toman valor en esta característica.
- En etiquetas compuestas si al menos uno de sus términos está en otro idioma, se considerará como etiqueta en otro idioma.

2.2. Neologismo: al no haber sido posible contar con lingüistas y profesionales de cada una de las áreas disciplinares de los TFMs, para la detección de neologismo se han tenido en cuenta los siguientes criterios:

- Si la etiqueta ha aparecido en fecha reciente.
- Si la etiqueta no parece registrada en el Diccionario de la Real Academia Española u otros diccionarios.

2.3. Errores ortográficos en las etiquetas asignadas por el autor al TFM

- Si las etiquetas solo presentan el error ortográfico en los términos asignados por el autor, y en el resto del trabajo sí están correctamente escritas, tomarán valor en esta característica, pero en el resto la etiqueta se analizarán sin tener en cuenta dicho error ortográfico. En el caso en que la etiqueta con error ortográfico esté presente en todo el texto, no tomará valor en la variable de coincidencia exacta.

2.4. Número gramatical: se analiza si la etiqueta es asignada en plural o singular.

- Etiquetas consideradas anteriormente como nombres propios, acrónimos y etiquetas en otro idioma diferente al castellano, no se tendrán en cuenta para el análisis de esta variable.

3. COINCIDENCIA ETIQUETA-TEXTO

3.1. Coincidencia exacta en el TFM: se analiza la coincidencia de la etiqueta durante el desarrollo del TFM.

- Se han considerado como coincidencia exacta:
 - Cuando la etiqueta aparece en el TFM en plural o singular.
Ejemplo: Palabra clave asignada: *populismos*
En el texto: *populismo*
 - Cuando a la etiqueta compuesta se le añade algún conector (preposiciones, determinantes, etc.) o viceversa.
Ejemplo:
Palabra clave asignada: *empleadas hogar/ política exterior brasileña*
En el texto: *empleadas del hogar/ política exterior en Brasil*
- Para que las etiquetas compuestas puedan considerarse como coincidencia exacta, deben aparecer en el texto con todos sus términos. Es decir, en el caso de que uno o varios términos de la etiqueta no estén presentes, esta etiqueta no tomará valor en esta característica.
 - En etiquetas compuestas formadas por un acrónimo seguido del desarrollo de sus siglas: si en el texto solo aparece el término desarrollado, sin ir acompañado del acrónimo o viceversa, la etiqueta no se considera coincidencia exacta.

3.2. Variación por sinónimos: se analiza la presencia de sinónimos de la etiqueta durante el trabajo.

- Por operatividad la búsqueda de sinónimos de etiquetas se ha realizado principalmente en los apartados de resumen e introducción de los trabajos. Si estos apartados son muy reducidos se ha pasado al marco teórico u otro apartado. En otras ocasiones, cuando se tenía claro conocimiento de sinónimos de la etiqueta, se han intentado localizar dentro del texto con la opción de búsqueda interna.
- Las etiquetas compuestas han sido consideradas como sinónimos cuando al menos uno de los términos que forman la etiqueta tenían un sinónimo.

Ejemplo: *Empresas mexicanas / compañías mexicanas*
Trabajadoras domésticas/empleadas del hogar

3.3. Variación morfológica: variación del morfema de la etiqueta (por género, número, prefijo, sufijo, aumentativo, diminutivo y otras terminaciones) manteniendo la raíz/lexema. La búsqueda de las etiquetas con variación morfológica se ha realizado del mismo modo que en la anterior variable.

- En etiquetas compuestas, han sido consideradas como variación morfológica cuando al menos uno de los términos que forman la etiqueta han sufrido algún tipo de variación.

Ejemplo: *Cursos Online/Curso Online*.

4. UBICACIÓN DE LA ETIQUETA

Se analiza la presencia de la etiquetas en los siguientes apartados:

4.1. Título

4.2. Resumen

4.3. Cuerpo

4.4. Solo Palabras Clave: las etiquetas que solo están presentes en esta variable, no toman valor en la variable de "Coincidencia exacta".

- Las etiquetas que aparecen en el título, resumen o cuerpo con diferente número gramatical (singular o plural) del de la etiqueta asignada, o se le añade algún conector (preposiciones, determinantes, etc.), también han sido contabilizadas en este criterio.

Ejemplo: En Resumen: *Dimensión latente*. Palabra clave: *Dimensiones latentes*

En el texto: *Política exterior en Brasil*. Palabra clave: *política Exterior brasileña*

5. CONTENIDO DE LA ETIQUETA

Esta variable es excluyente o dicotómica.

5.1. Materia: etiquetas/palabras clave relacionadas con la materia y temática del TFM.

5.2. Metodología: etiquetas/palabras clave relacionadas con la metodología llevada a cabo en el TFM.

6. COINCIDENCIA EN EL REPOSITORIO GREDOS

Se analiza la coincidencia entre las palabras clave de autor ubicadas en el TFM, y las que están en el registro de GREDOS.

- Cuando las etiquetas de autor y las de GREDOS tengan distinto número gramatical (plural/singular), serán consideradas como coincidencia exacta.

7. NÚMERO DE ETIQUETAS POR TFM: se analiza el promedio de etiquetas por TFM.

3.2.2 Población y muestra

La población de análisis está compuesta por los 527 TFM's ubicados en GREDOS, de los que se obtuvo una muestra final de 187 TFM's (35,48%), esta reducción de la muestra se debe a la ausencia de trabajos con palabras clave de autor. De estos 187 trabajos se han recopilado y analizado un total de 963 etiquetas, donde no se ha realizado ningún muestreo pues se han utilizado la totalidad de palabras clave para el análisis.

Destacar que el número de trabajos analizados ha ido incrementándose progresivamente, ya que existe una periódica incorporación de TFM's en GREDOS. Por este motivo, se han analizado los trabajos ubicados en el repositorio hasta el 25/04/2017.

3.2.3 Métodos de análisis

3.2.3.1 Recopilación de datos

El primer paso llevado a cabo para comenzar con el estudio ha sido localizar y recopilar los TFM's y palabras clave dentro del repositorio. Para acceder a la colección se ha partido de la comunidad "Repositorio Docente" donde se ubica la subcomunidad "Trabajos de Fin de Máster". Una vez dentro del listado de másters impartidos en la USAL, se ha utilizado el filtro "Fecha de publicación descendiente", y se ha ido accediendo de forma individual a cada uno de ellos para comprobar si disponían de palabras clave. Era necesario realizar esta tarea ya que en el caso de que el trabajo no disponga de palabras clave, el repositorio añade las suyas propias al registro del TFM, por lo que era obligatorio hacer esta comprobación para asegurarnos de recuperar las palabras clave de autor y no los términos controlados añadidos por GREDOS.

En caso afirmativo, se han extraído las palabras clave y otros datos relevantes del TFM. Para el almacenamiento, tratamiento y representación de datos se ha utilizado el programa de hojas de cálculo Microsoft Excel 2010. Posteriormente, se llevó a cabo un análisis de los datos extraídos, resultado del estudio de las anteriores variables presentadas. Como se puede observar en la imagen 4, la distribución utilizada ha sido la siguiente:

- Las columnas A, B, C y D se han utilizado para ofrecer la información básica del trabajo analizado, como son: el nombre del máster del TFM, la rama a la que pertenece el máster, el título del trabajo analizado y las palabras claves del mismo. Se han utilizado tantas filas en la columna "etiqueta"(D), como palabras clave había en el trabajo.
- Las siguientes columnas se han utilizado para ubicar las variables a analizar.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1					ETIQUETA SEGÚN EL Nº DE TÉRMINOS								
2					Tipo de etiqueta SIMPLE (Punto de vista lingüístico)								
3	Máster	Rama	TFM	Etiqueta	Etiqueta simple	Sustantivo	Adverbio	Adjetivo	Número	Acrónimo	Verbo	Nombre Propio	Etiqueta compuesta
4													
5													
6													
7													
8													
9													

Imagen 13. Distribución en la hoja de cálculo utilizada para el análisis de las palabras clave

La escala establecida de valores que puede tomar cada etiqueta/palabra clave, en cada variable son:

VALOR 0

Este valor solo es aplicable en la variable “Número gramatical”, y se asignará a las etiquetas en singular. Las plurales obtendrán un valor 1.

VALOR 1

Este valor es asignado cuando existe una coincidencia exacta entre la etiqueta y la variable analizada. En el caso de que la etiqueta no cumpla o coincida con una variable determinada, no se asignará ningún valor en la celda correspondiente, por lo que toda casilla vacía indica la no coincidencia entre etiqueta y variable. Por ejemplo, si la etiqueta es “Barcelona”, esta tomará valor 1 en la variable “Nombre Propio”, y en “Etiqueta compuesta” no obtendrá ningún valor.

VALORES

1. 2. 3. 4. 5...

Solo obtendrán estos valores (1 y >1) las etiquetas compuestas. Indican el número de términos que forman la etiqueta. Por ejemplo: la etiqueta “Biblioteca Digital”, tomará un valor “2” en “Etiqueta compuesta”.

Durante el análisis, ha sido necesario y esencial trabajar directamente con el TFM. Por operatividad, el estudio individual de etiquetas en criterios como “Variación morfológica” y “Variación por sinónimos”, se ha efectuado principalmente en los apartados de resumen, introducción y marco teórico, u otro apartado en el caso de brevedad de los anteriores. Es cierto que también se ha utilizado la opción de búsqueda interna del texto, cuando por ejemplo, se tenía conocimiento de sinónimos de la etiqueta, o buscando la raíz de la palabra clave para detectar cualquier variación morfológica.

3.2.3.2 Tratamiento y análisis de datos

Una vez recopilados todos los TFMs junto con sus palabras clave, se lleva a cabo recuento de los valores de cada variable, y el cálculo de sus porcentajes en la hoja de cálculo Excel para la elaboración de gráficos y la posterior interpretación y explicación de los resultados obtenidos.

3.2.4 Limitaciones y problemas encontrados

- Una de las limitaciones iniciales del trabajo ha sido la de detectar una comunidad que tuviera palabras clave asignadas por el autor. Una vez que se localizó la comunidad “Trabajos de Fin de Máster”, durante el análisis se encontró con un alto número de trabajos que no tenían asignadas palabras clave, por lo que la población de análisis quedó reducida. Es cierto, que en estos casos GREDOS asigna las suyas utilizando vocabularios controlados, por lo que estos TFMs quedaron finalmente descartados para ser analizados.
- Otra cuestión relacionada con GREDOS, fue que al entrar en los registros de cada trabajo se observó como el repositorio no respetaba en ocasiones las palabras clave asignadas por el autor. De modo que este detalle se utilizó para estudiar la coincidencia de las palabras clave de autor con las asignadas en su registro en GREDOS.
- La legislación de la USAL permite la presentación de trabajos en idiomas diferentes al castellano. Durante el análisis se encontraron TFMs en portugués e inglés. Dado que el número de trabajos no es demasiado elevado, no se han descartado para el análisis. Su estudio se ha centrado en los apartados “resumen” y “palabras clave” en castellano que incluían dichos trabajos, y por lo tanto, en la categoría “Ubicación de la etiqueta”, las palabras clave solo podrán tomar valor en “título”, “resumen” o “solo palabras clave”, y no en “cuerpo”.
- En un principio se comenzó a analizar la frecuencia aparición de la etiquetas, pero debido a su dificultad decidió eliminarse. Para obtener la frecuencia se utilizó la opción de búsqueda interna en el texto, y se obtuvieron etiquetas que no deberían ser contabilizadas, como las que aparecían en referencias bibliográficas, o en el encabezado de página (cuando la etiqueta estaba en el título), por lo que estos resultados no eran válidos.
- Otro error que se encontró en GREDOS fue la existencia de trabajos de fin de grado dentro del listado de TFMs. En especial en el Máster Oficial en Sistemas de Información Digital. Estos trabajos no fueron contabilizados ni analizados.

4 Resultados y discusión

El análisis realizado a la totalidad de palabras clave/etiquetas de TFM, disponibles en el repositorio GREDOS, ha permitido la obtención de resultados convenientes dentro de los objetivos perseguidos. El proceso de clasificación de las etiquetas, y su posterior recuento y aplicación de los cálculos correspondientes, han proporcionado los resultados que se presentan a continuación. También se realiza una comparación con otros estudios de interés, que en la medida de lo posible, permiten establecer una comparación con los resultados obtenidos en el presente trabajo.

El total de Trabajos de Fin de Máster depositados en GREDOS (a fecha de 25 de abril de 2017) es de 527, de los cuales solo 187 (35,48%) trabajos disponían de palabras clave. El número total de etiquetas analizadas ha sido 963.

A pesar de los resultados favorables del primer muestreo realizado para observar si los TFM de diferentes másters disponían de palabras claves, el resultado final no fue el esperado, ya que solo el 35,48% de los trabajos en GREDOS son descritos con palabras clave de autor. Al contrario ocurrió en el estudio de Tomás-Casterá et al. (2009), que al estudiar las palabras clave utilizadas en los artículos de las revistas especializadas en nutrición indizadas en la Scientific Electronic Library Online (Red SciELO), para compararlas con los descriptores de Ciencias de la Salud (DeCS), solo el 14,24% de los artículos de revistas no presentaron palabras clave.

Ramas de másters de TFM con palabras clave

La primera variable analizada ha sido “Rama de másters con palabras clave” impartidos en la USAL. Como se muestra en el gráfico 1, la rama con mayor número de TFM etiquetados es “Ciencias Sociales y Jurídicas” con 115 trabajos (61,50%).

En segunda y tercera posición se encuentra con valores muy similares las ramas “Ciencias” y “Ciencias de la salud” con 29 (15,51%) y 28 (14,97%) trabajos respectivamente. Como se puede observar estos valores están muy alejados de la primera rama. Por último, los resultados más bajo los obtienen las ramas de “Arte y humanidades” con total de 11 trabajos e “Ingeniería y Arquitectura” con el porcentaje mínimo, ya que solo el 2,14% de la muestra, es decir, 4 trabajos son etiquetados por sus autores.

Que la rama de “Ciencias sociales y jurídicas” sea con diferencia la que mayor porcentaje alcanza, podría ser debido a que dentro de esta disciplina el etiquetado de información es objeto de estudio y por ello es ampliamente utilizado. Pero en este contexto, la causa de estos resultados se debe a decisión de cada comisión académica de másters, que son las que establecen los criterios para la realización y depósito del TFM, decidiendo si los autores deben o no añadir palabras clave a sus trabajos.

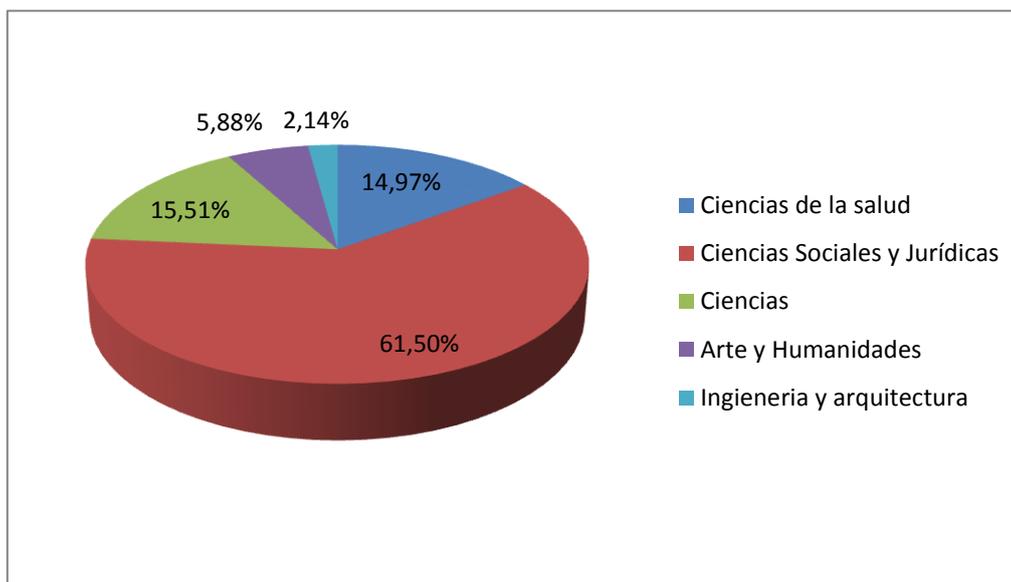


Gráfico 1. Distribución porcentual de ramas de máster con TFM con palabras clave.

Ramas de másters con palabras clave	Total TFM	%
Ciencias de la salud	28	14,97%
Ciencias Sociales y Jurídicas	115	61,50%
Ciencias	29	15,51%
Arte y humanidades	11	5,88%
Ingeniería y arquitectura	4	2,14%
Total TFMs	187	100,00%

Tabla 2. Ramas de másters con TFM con palabras clave de autor

Estos resultados tan dispares podrían normalizarse si el Reglamento de Trabajos Fin de Máster de la Universidad de Salamanca estableciera un único modelo de presentación y depósito de TFM y no delegará esta tarea con total libertad a cada comisión académica de máster, ya que como se expone anteriormente solo el 35,48% de los TFM ubicados en GREDOS disponen de palabras clave.

Etiqueta según el número de términos que la forman

Como se observa en el gráfico 2, los resultados obtenidos no están muy alejados entre sí, ya que de las 936 etiquetas de la muestra, 503 (52,23%) son etiquetas simples y 460 (47,77%) son compuestas. Por lo que los autores/etiquetadores, no presentan una tendencia destacable a la hora de utilizar etiquetas simples o compuestas, aunque sí es cierto que existe una leve preferencia por las simples.

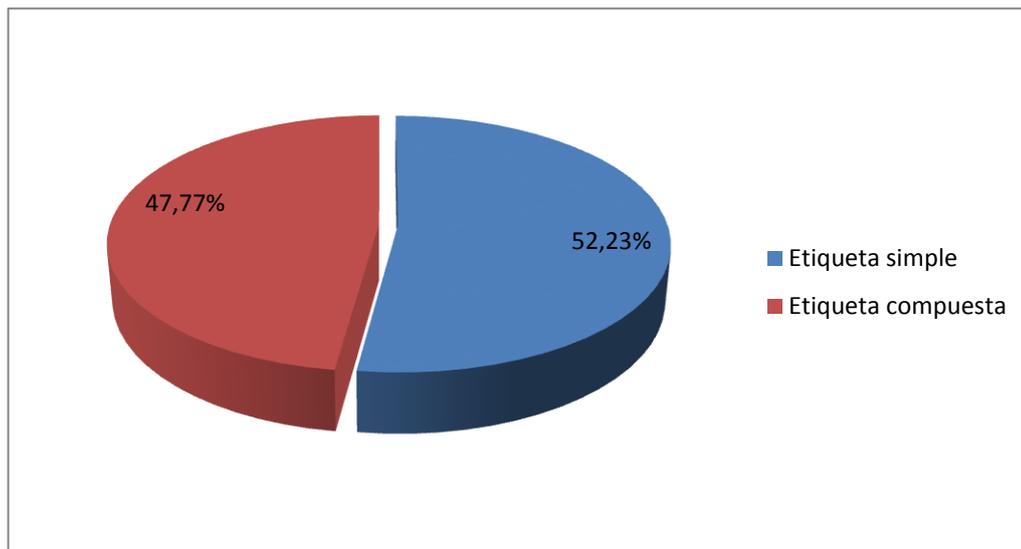


Gráfico 2. Distribución porcentual de la aparición de etiquetas simples y compuestas

Otro estudio que analizó esta variable fue el realizado por Spiteri (2007), quien recopiló un conjunto de etiquetas de los marcadores sociales Delicious y Furl, y del motor de búsqueda de blogs Technorati, consiguiendo unos resultados con valores altos respecto al uso de etiquetas formadas por un solo término. En Delicious infiere un porcentaje de uso de etiquetas simples del 93%, en Furl un 76%, y un 80% en Technorati. El resultado que obtuvo en Delicious fue debido a que no permite el uso de espacios entre los diferentes términos de la etiqueta. Independientemente a esto, las otras dos plataformas alcanzaron unos valores de uso muy elevados. Igualmente, en el trabajo de Heckner et al. (2008), donde analizaron un conjunto de etiquetas utilizadas en Connotea, las etiquetas formadas por más de un término no superaron el 29,13% de uso.

Recientemente, Choi & Syn (2015) analizaron un conjunto de etiquetas extraídas del portal digital NINES, y obtuvieron que el 75,90% de las etiquetas estaban formadas por un único término y el resto (24,09%) estaban formadas por más de uno.

En conclusión, aunque en nuestro estudio se han obtenido resultados semejantes para ambas tipologías, los trabajos de Spiteri (2007), Heckner et al. (2008) y Choi & Syn (2015) sí presentan una tendencia clara hacia la descripción de información con etiquetas simples. Una de las razones, podría ser debido a la facilidad de asignar estas etiquetas frente a la complejidad de crear etiquetas compuestas donde el usuario, por ejemplo, debe elegir el sistema de separación entre términos, número gramatical, orden de los términos, tiempo verbal, etc.

- **Tipo de etiqueta simple (análisis gramatical)**

Dentro de la tipología de etiqueta simple (503 etiquetas), se han establecido 7 variables: nombre común, adverbio, adjetivo, número, acrónimo, verbo y nombre propio, resumidas en la siguiente gráfica.

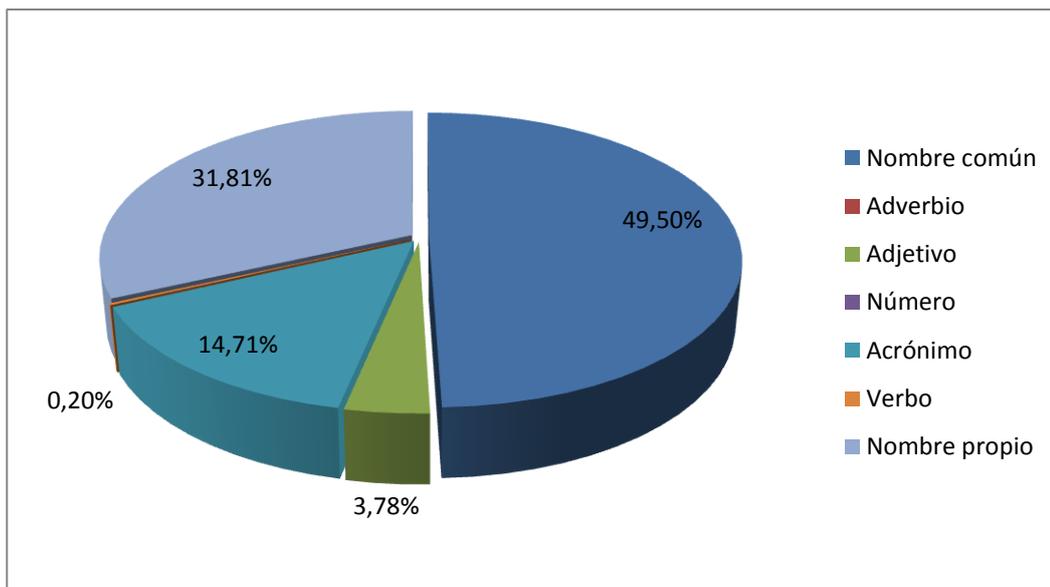


Gráfico 3. Distribución porcentual de los tipos de etiquetas simples (análisis gramatical)

Tipo de etiqueta simple	Total etiquetas	%
Nombre común	249	49,50%
Adverbio	0	0%
Adjetivo	19	3,78%
Número	0	0%
Acrónimo	74	14,71%
Verbo	1	0,20%
Nombre propio	160	31,81%
Total	503	100,00%

Tabla 3. Resultados de los tipos de etiquetas simples (análisis gramatical)

Como era de esperar, la variable “Nombre común” es la más frecuente y utilizada para la descripción de información mediante palabras clave, ya que recopila un total de 249 etiquetas con un porcentaje de uso del 49,50%, representando casi la mitad del total de la muestra de etiquetas simples. Seguido se encuentran las variables “Nombre propio” y “Acrónimo”, con 160 (31,81%) y 74 (14,71%) etiquetas respectivamente. Estas son las variables con los resultados más representativos dentro de esta tipología.

Los resultados más bajos los han obtenidos las variables “Adjetivo” con 19 etiquetas y un 3,78%, y “Verbo” presente solo una vez y representando solo el 0,22% de la muestra. De las variables “Adverbio” y “Número” no se obtuvieron resultados.

En relación con el estudio de Heckner et al. (2008), los resultados son en general bastantes parecidos. Los nombres comunes y nombres propios los agruparon en una sola variable, alcanzando un total del 72% de la muestra. En el caso de nuestro trabajo, si unimos las variables “Nombre común” y “Nombre propio” obtendríamos un porcentaje del 81,31%. En ambos estudios, el porcentaje de etiquetas afectadas por esta variable es elevado y representan la mayor parte de la muestra.

Una tipología con resultados casi idénticos en ambos estudios es “Acrónimo”, donde Heckner et al. (2008) obtuvo un 15% y el presente trabajo alcanza un 14,71%. De igual forma sucede con las variables “Número” con un 1% y “Verbo” y “Adverbio” ambas con una sola etiqueta, siendo valores irrelevantes en la muestra. Sin embargo, el presente estudio solo obtiene un porcentaje del 3,78% en la variable “Adjetivo”, frente al 12% que alcanzan los anteriores autores.

Otro estudio que analizó las características lingüísticas y gramaticales de etiquetas fue Spiteri (2007). En la tabla 4 se puede muestra los resultados que obtuvo en las diferentes plataformas analizadas:

	Delicious	Furl	Technorati
Sustantivos	88%	71%	86%
Adjetivos	6%	6%	3%
Adverbios	0%	0%	0%

Tabla 4. Resultados de los tipos de etiquetas simples (análisis gramatical). Resultados obtenidos por Spiteri (2007).

Como se puede observar, en las tres plataformas analizadas el porcentaje de sustantivos es elevado, representando más de la mitad de la muestra en todas las plataformas. No se especifica si en los sustantivos se recogen también nombres propios, aunque el autor los analizó en otro apartado con la variable “entidades únicas”. Además este estudio analizó dentro de esta clasificación variables sobre frases nominales y sustantivos verbales no utilizados en esta comparación. Los resultados de este estudio son similares a los obtenidos en nuestro estudio (81,31%, formado por nombres comunes más nombres propios) y al de Heckner et al. (2008) (72%, formados por nombres comunes más nombres propios).

Si se compara de forma individual los resultados obtenidos en la variable “nombres propios/entidades únicas”, también se alcanzan unos resultados muy igualados, ya que nosotros alcanzamos un porcentaje del 31,81%, frente al 22% de etiquetas presentes en Delicious, 14% en Furl y 49% en Technorati.

En relación con el uso de adjetivos (6% en Delicious y Furl, y 3% en Tecnorati) y de adverbios (0%), los resultados siguen mostrando una presencia muy escasa o inexistente en estas tipologías, al igual que en el presente trabajo y en el de Heckner et al. (2008). Normalmente cuando se utilizan adjetivos y adverbios para describir contenidos, el usuario etiqueta de una forma egoísta o excesivamente personal, por lo que estos resultados son favorables ante esta cuestión.

Spiteri (2007) también analizó las etiquetas acrónimos y abreviatura, pero en una categoría aparte ya que existen acrónimos que también los valoró como nombres propios (al contrario que en el trabajo de Heckner et al. (2008) y al nuestro, donde se han analizado de forma separada). Aun así, los resultados que se obtuvieron no son tan dispares a los obtenidos en el presente estudio, ya que los acrónimos y abreviaturas representan el 22% de etiquetas en Delicious, 16% en Furl, y 19% en Technorati, frente a nuestro 14,71%.

Continuando con el análisis de estas variables, algunos autores afirman que el uso de acrónimos crea ruido documental en sistemas y unidades de información, ya que

existen usuarios que desarrollan el acrónimo para asignarlo como etiquetas, mientras que otros utilizan solamente el acrónimo. Esto da lugar a que una única etiqueta sea representada de diferentes formas (Soler Monreal & Gil Leiva, 2010). En el presente estudio se ha observado como la etiqueta “Tecnologías de la Información y Comunicación” se utiliza de tres formas diferentes:

1. Tecnologías de la información y comunicación
2. TIC
3. TIC (Tecnologías de la Información y Comunicación)

Además, en las etiquetas donde se utilizan el acrónimo más sus términos desarrollados, existe también el inconveniente del orden y la estructura en la creación. Por ejemplo, siguiendo con el anterior ejemplo, ciertos trabajos añaden el acrónimo indistintamente tanto antes como después del conjunto de términos que forman la etiqueta. Ejemplo: “Tecnologías de la Información y Comunicación (TIC)” y “(TIC) Tecnologías de la Información y Comunicación”. Otro problema observado en el uso de acrónimos es la utilización de mayúsculas y minúsculas, por ejemplo, la etiqueta “CHAID” aparece también como “chaid”.

Estos problemas hacen necesario la creación de indicaciones o recomendaciones que establezcan una estructura fija para la creación de este tipo de etiquetas, unificando todas las variedades posibles en la asignación de palabras clave e intentando presentarlas con la mayor claridad para que sean aceptadas por toda la comunidad de usuarios.

En definitiva, los resultados que se obtuvieron en los trabajos de Spiteri (2007) y Heckner et al. (2008), se ajustan a los obtenidos en el presente estudio. Y se destaca el uso de sustantivos (agrupando nombres propios y nombres comunes), seguido de acrónimos, y finalmente la ausencia o uso mínimo de adverbios, adjetivos, números y verbos.

- **Etiqueta compuesta (Nº de términos que la forman)**

Se han analizado las 460 etiquetas compuestas, en función del número de términos que la forman.

Número de términos por etiqueta	Ocurrencias	%
2	387	84,13%
3	65	14,13%
4	5	1,09%
5	2	0,44%
6	0	0%
7	1	0,20%
Total	460	100,00%

Tabla 5. Total de etiquetas en función del número de términos que la forman

Como muestra la tabla 5, de las 460 etiquetas compuestas un 84,13%, es decir, 387 están formadas por dos términos. A continuación se ubican las de tres términos alcanzando un total de 65 etiquetas (14,13%). Estos tipos de etiquetas compuestas, con dos y tres términos son las más concurrentes, aunque sin lugar a dudas, las etiquetas formadas por dos términos, encabezan con un resultado muy por encima del resto. Finalmente en los resultados más bajos, se encuentran cinco etiquetas formadas con cuatros términos, dos etiquetas de cinco términos y una etiqueta de siete términos.

Otros autores que analiza esta cuestión, son Heckner et al. (2008), donde al igual que el presente trabajo, las etiquetas formadas por dos términos, son las que mayor presencia tienen con un 83,29% de las 347 etiquetas compuestas totales. Las formadas por tres términos, también obtienen un resultado similar al obtenido en nuestro estudio, pues consiguen el 13,26%. Seguidamente, se encuentran siete etiquetas formadas por cuatro términos, dos etiquetas por cinco términos, una etiqueta con seis términos y dos etiquetas con ocho términos.

Por lo tanto, el autor tiende a etiquetar con etiquetas compuestas de dos términos (además de las etiquetas simples), puesto que son términos más sencillos y menos complejos a la hora de su construcción, que una etiqueta formada por 6, 7 o incluso 8 términos. Finalmente, aunque en nuestro estudio se han obtenido resultados semejantes en la utilización de etiquetas simples y compuestas, los trabajos de Spiteri (2007) y Heckner et al. (2008) sí presentan una tendencia clara de los autores a utilizar etiquetas simples.

Las etiquetas compuestas presentan una mayor complejidad a la hora de su construcción, ya que al no haber unas indicaciones establecidas surgen errores a la hora de elegir un sistema de separación entre términos, lo que da lugar a que una misma etiqueta aparezca escrita de diferentes formas (ejemplo: “Análisis de Clúster” y “Análisis Clúster” o “HJ-BILOT” y “HJ Biplot”), o a etiquetas muy extensas (ejemplo: “Informe Final del Grupo Incubador de Datos Vinculados de Bibliotecas”). Por lo que las palabras clave con más de un término son un problema en la indización libre y deben controlarse aplicando normas para su creación.

Clasificación lingüística

Dentro de esta tipología, existen tres variables analizadas:

- **Etiqueta en otro idioma (diferente al castellano)**

De las 963 etiquetas de la muestra 100 (10,38%) son asignadas en otro idioma diferente al castellano. Independientemente de los nombres propios y acrónimos en inglés, se han identificado etiquetas en inglés y portugués, pues la legislación de la USAL permite la realización de TFMS en ambos idiomas. Aunque no es un resultado elevado, esta cuestión podría resolverse estableciendo unas normas donde se obliguen a los autores a añadir palabras clave en castellano, independiente del idioma en el que se desarrolle posteriormente el trabajo. No obstante existe una gran cantidad de TFMs que aunque han sido redactados en otro idioma, han añadido un resumen y palabras clave en castellano, por lo que se han utilizado esas etiquetas para el análisis, pero también, ha habido autores que solo las han asignado en inglés o portugués, y estas son las que forman una parte del 10,38% de la muestra.

Esta cuestión puede variar en función del idioma que predomine en las plataformas de análisis. Guy & Tokin (2006) al analizar Flickr y Delicious, observaron como las etiquetas en inglés alcanzaron unos valores altos, de 46% en Flickr y 64% en Delicious, además de detectar 5 idiomas más y un grupo de etiquetas en idiomas desconocidos. Otra plataforma donde sucedió lo mismo fue Connotea, donde solo el 1,51% de las etiquetas estaban en español e italiano y un 98,49% en inglés (Heckner et al., 2008). En conclusión, tanto en nuestro estudio como en los dos anteriores existe un idioma que predomina notablemente en la asignación de etiquetas, seguido de un porcentaje menor formado por otros idiomas menos utilizados.

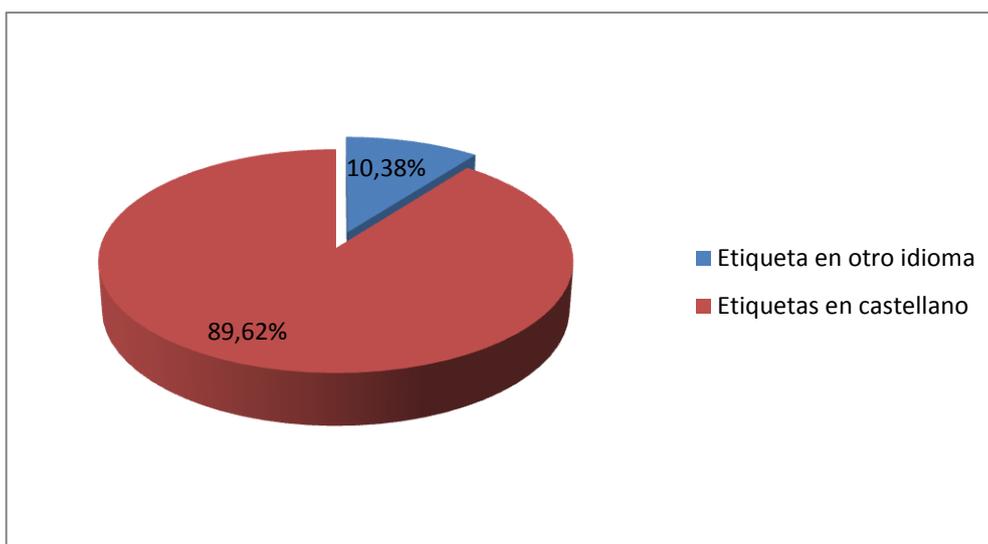


Gráfico 4. Distribución porcentual de la aparición de etiquetas en otro idioma diferente al castellano

Los resultados obtenidos no son alarmantes, ya que además GREDOS ha incorporado a los registros la traducción al castellano de dichas etiquetas. Pero hay que tener presente que la variedad de idiomas es uno de los motivos del aumento de la ambigüedad en unidades de información, que complican la tarea de recuperación de información a usuarios que no tienen conocimiento del idioma y no disponen de un vocabulario controlado multilingüe.

- Neologismos

De las 963 etiquetas de la muestra, 30 (3,12%) toman valor en esta variable (grafico 5). Recordemos que estas etiquetas representan conceptos que han aparecido recientemente y aun no son aceptados por los diccionarios. Normalmente el usuario recurre a estos conceptos cuando no dispone de otra alternativa y el contenido del recurso que va a describir la etiqueta es actual y novedoso. El resultado obtenido es bajo y no constituye una proporción relevante en este trabajo, además estas etiquetas suelen hacer referencia a cuestiones informáticas y tecnológicas o situaciones sociales (ejemplo: *big data*, *whatsapp*, *justicia 2.0*, *ciberbullying*).

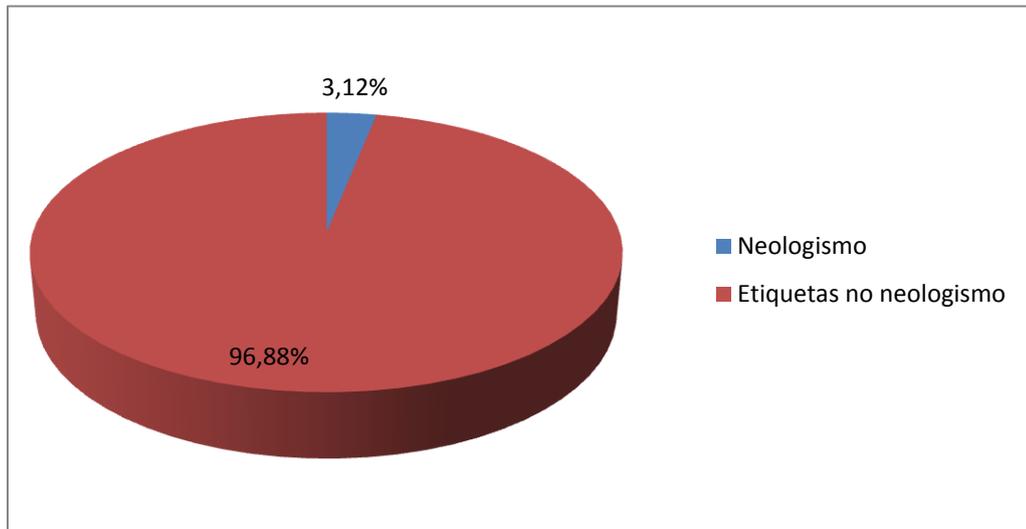


Gráfico 5. Distribución porcentual de las etiquetas neologismo

Esta tipología también ha sido estudiada en otras plataformas, siendo el porcentaje de neologismos en Delicious de 3%, en Furl de 10% y en Technorati de 6% (Spiteri, 2007). En el estudio de Heckner et al., (2008), solo el 0,83% de la muestra son valoradas como neologismos. Estos datos se ajustan a los obtenidos en el presente trabajo de 3,12% de la muestra.

Estos resultados tan bajos indican que los autores son bastante conservadores en la descripción, y prefieren utilizar etiquetas conocidas y comúnmente empleadas dentro de su comunidad de trabajo o área de estudio, además de ser aceptadas en diccionarios. Hay que tener en cuenta que los profesionales de la información no utilizarían estos términos en el proceso de indización si tuvieran que atenerse a vocabularios controlados o lenguajes documentales poco actualizados.

- Errores ortográficos

Se han obtenido 5 etiquetas con errores ortográficos, es decir, solo un 0,52% de la muestra analizada (gráfico 6). El riesgo de cometer errores existe al crear etiquetas en lenguaje natural y sin ningún tipo de supervisión. En este caso, el escaso valor obtenido es buena señal pues las faltas ortográficas entorpecen el proceso de recuperación de información dificultando el acceso a los recursos descritos por estas etiquetas erróneas.

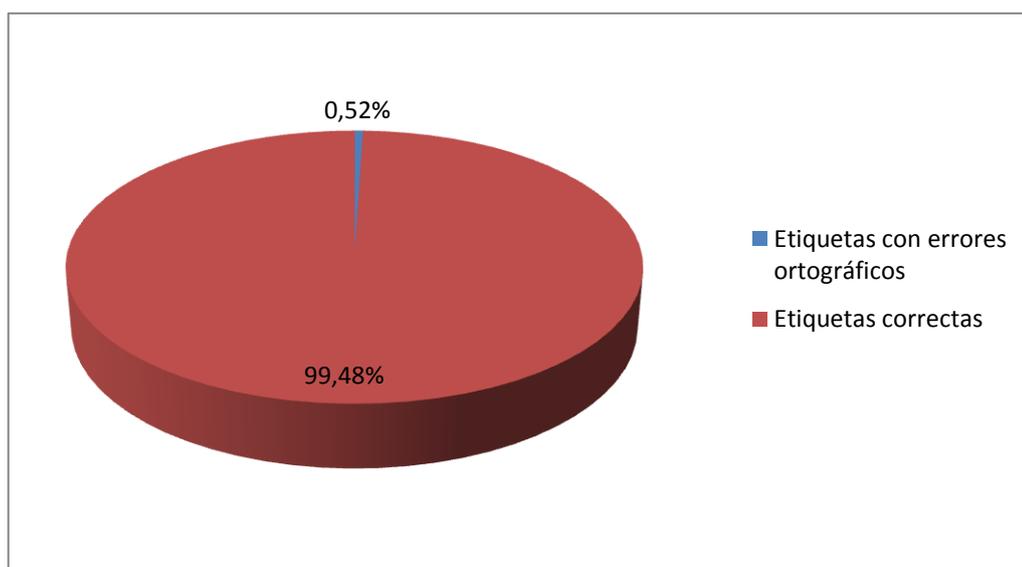


Gráfico 6. Distribución porcentual de las etiquetas con y sin errores ortográficos

En el trabajo de Guy & Tonkin (2006) el número de etiquetas afectadas por esta variable fue de un 40% en Flickr y un 28% en Delicious. Unos resultados que difieren con los obtenidos en este trabajo. Sin embargo, esta diferencia se explica a que estos autores agruparon dentro de esta tipología etiquetas que representaban expresiones compuestas multilingües y etiquetas que no fueron identificadas por el software, debido a un idioma desconocido o a la utilización de códigos y símbolos.

Posteriormente, en el trabajo de Spiteri (2007) el número de etiquetas mal escritas constituye un total de 4% de las etiquetas en Delicious, 3% en Furl, y 2% en Technorati. Heckner et al., (2008) también obtuvo un valor bajo con solo un 1,5% de etiquetas con errores ortográficos presentes en Connotea.

Esta tasa de error mínima indica que el autor es cuidadoso y presta atención a la hora de asignar las etiquetas a sus trabajos. En GREDOS no existe un sistema de recomendación de etiquetas que pueda orientar al usuario ante estos fallos, pero Heckner et al., (2008) sí expuso que estos resultados pueden atribuirse a que sistemas y marcadores sociales presentan posibles sugerencias creadas, mientras que el usuarios está escribiendo la etiqueta, a partir de otras ya asignadas anteriormente.

Se ha comparado también si GREDOS en sus registros añade las etiquetas manteniendo los errores o las corrige. De las cinco etiquetas erróneas, solo en uno de los casos el repositorio añade al registro la etiqueta escrita correctamente. Otras dos etiquetas no son corregidas por GREDOS, y las dos restantes no aparecen en los registros (el repositorio a agregado otras etiquetas nuevas).

Finalmente, en la indización en lenguaje natural siempre existirá la posibilidad de crear etiquetas con errores ortográficos, mientras que no exista ningún tipo de supervisión o cierto control que avise al usuario de la creación de una etiqueta con algún tipo de fallo. De esta forma no se perdería la esencia del etiquetado libre, ya que no afectaría a su construcción, ni al contenido de la etiqueta.

- Número gramatical

El número de etiquetas analizadas en esta variable es de 697 (excluyendo del análisis etiquetas nombre propio, acrónimos y etiquetas en otro idioma diferente al castellano). Si el total de la muestra es de 963, el 72,37% de las etiquetas se presentan en plural o singular.

De las 697 etiquetas, un 73,74% están en singular y el 26,26% en plural (gráfico 7). Los resultados manifiestan una preferencia clara por el uso del singular a la hora de asignar etiquetas.

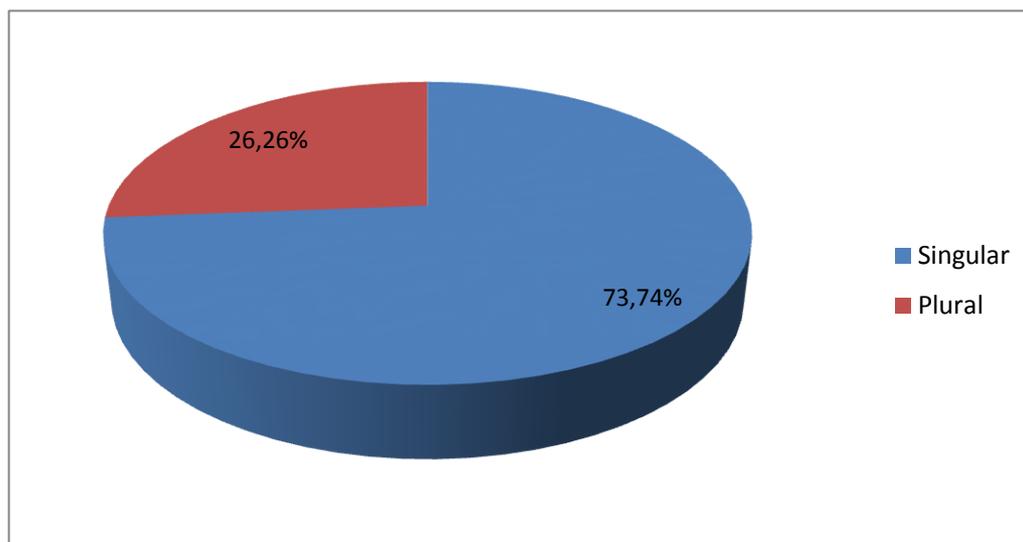


Gráfico 7. Distribución porcentual de la aparición de etiquetas en plural y singular

Guy & Tokin (2006) al analizar el número gramatical de las etiquetas también observaron una preferencia en el uso del singular, ya que en Flickr y en Delicious solo se alcanzaron el 8% y 11% de etiquetas en plural respectivamente. Sin embargo, Spiteri (2007) estudió esta cuestión en las etiquetas sustantivos y obtuvo unos resultados variados, ya que la presencia de etiquetas en plural afectó a la muestra con unos porcentajes del 36% en Delicious, 34% en Technorati y a un 62% en Furl.

Aun así, dos cuestiones quedan claras: que el uso del singular predomina en la creación y asignación de etiquetas, y que la normalización del número gramatical de las etiquetas es necesario. La inconsistencia entre las formas plural y singular, los errores ortográficos y la presencia de etiquetas en varios idiomas, son los principales problemas del etiquetado libre (Guy & Tonkin, 2006).

Coincidencia de la etiqueta en el TFM

En esta tipología existen tres variables que analizan la presencia y las diferentes formas de aparición de las etiquetas dentro de los TFMs que describen.

- Coincidencia exacta de la etiqueta en el TFM

El número de etiquetas afectadas por esta variable es de 909. Dado que el total de la muestra es de 963, el 94,39% de las etiquetas están presentes durante el desarrollo de sus respectivos TFM (gráfico 8).

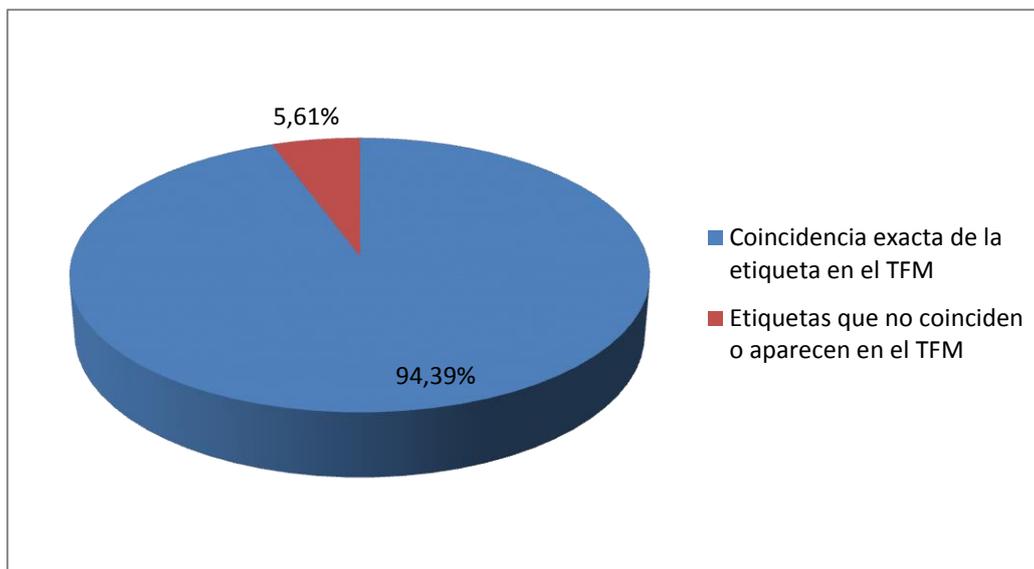


Gráfico 8. Distribución porcentual de las etiquetas que coinciden en el texto de sus respectivos TFM

En esta variable se utilizaron unos criterios metodológicos flexibles, en el sentido de que las etiquetas que aparecían en el trabajo con diferente número gramatical y que añadían algún conector (preposiciones, determinantes, etc.) o viceversa, se consideraron como coincidencia exacta. En el caso de etiquetas compuestas si debían aparecer la totalidad de los términos que forman la etiquetas.

El resultado obtenido es positivo ya que solo el 5,61% de las etiquetas o no están presentes de forma exacta (sobre todo en etiquetas compuestas) o no aparecen durante el desarrollo del TFM.

Se ha observado si las etiquetas que no presentan coincidencia en el texto, posteriormente toman valor como variación morfología (cambio del morfema manteniendo la raíz), y son mínimos los casos. Por lo que este pequeño porcentaje puede indicar que existen autores que utilizan palabras clave que no llegan a identificar correctamente (por ser más generales) o no concuerdan con el contenido su trabajo, ya que posteriormente no son utilizadas dentro del mismo. También se ha observado que en la mayor parte de las etiquetas que no coinciden en el texto, tampoco cuentan con sinónimos que puedan “suplir” esta ausencia.

En conclusión, existe un conjunto reducido de autores que no etiquetan de forma correcta sus trabajos. Esta cuestión podría resolverse mediante una alfabetización de los etiquetadores, estableciendo ciertas pautas generales para que el autor asigne etiquetas que represente correctamente el contenidos del recurso y evite la utilización de términos que entorpezcan la descripción y recuperación de información (Saavedra Saldívar, 2010).

- Variación por sinónimos

Se han obtenido 179 etiquetas que disponen de conceptos sinónimos dentro de los TFM's que describen. Si el total etiquetas analizadas son 963, se obtiene un porcentaje de 18,59% de etiquetas con sinónimos (gráfico 9).

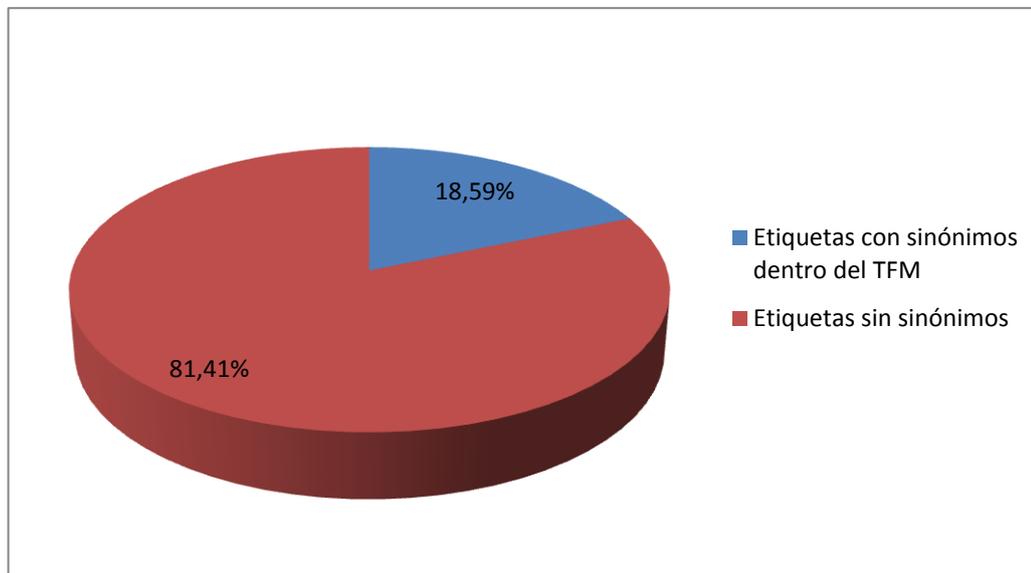


Gráfico 9. Distribución porcentual de las etiquetas con y sin sinónimos dentro del TFM

Siempre que la etiqueta esté presente en el texto, la aparición de sinónimos es positiva ya enriquece el contenido del TFM y facilita la comprensión del usuario/lector. Por lo que este valor obtenido es favorable aunque no destaque en la muestra analizada, ya que no alcanza el 20%.

- Variación morfológica

Se han obtenido 398 etiquetas que varían morfológicamente (por género, número, prefijos, sufijos, etc.) dentro de los TFM's (gráfico 10).

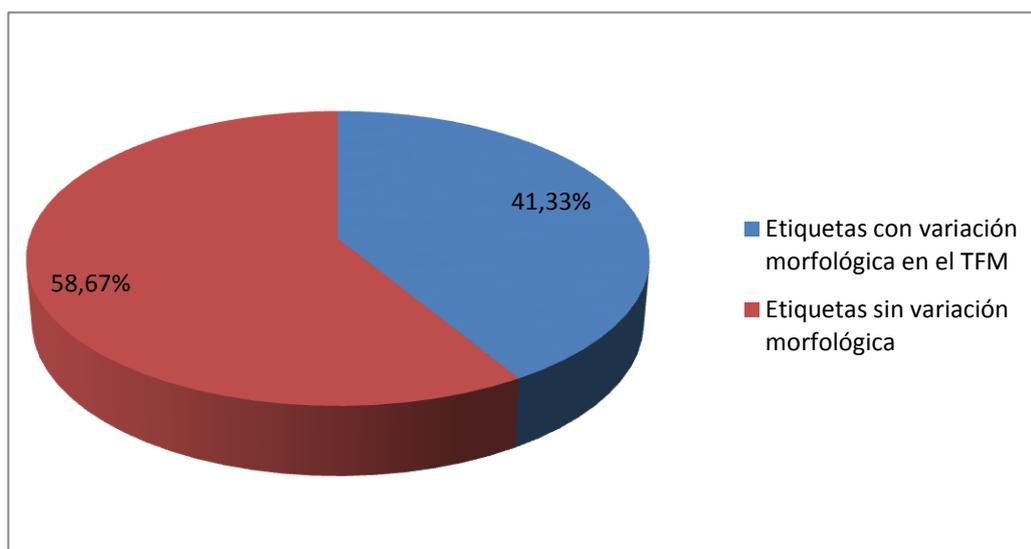


Gráfico 10. Distribución porcentual de las etiquetas con o sin variación morfológica dentro del TFM.

El 41,33% de las etiquetas que están presentes en el TFM varían morfológicamente, por lo que el concepto que representa la etiqueta es ampliamente utilizado en el trabajo. Tanto la etiqueta como sus diferentes formas, son parte de una misma familia léxica (Ejemplo: educar, educativo/a, contraeducación, educador/a, etc.).

El 58,67% de la muestra puede estar formada por nombres propios o acrónimos, ya que son mínimas las etiquetas de estas tipologías que presentan algún tipo de variación (Ejemplos: SICAV, SICAVs o Brasil, brasileño).

De forma general los resultados de estas tres variables se pueden interpretar como que los autores de los TFMs utilizan etiquetas de conceptos ya presentes dentro los trabajos de forma exacta. Que estos no suelen utilizar etiquetas que cuenten con sinónimos durante el desarrollo del mismo, ya que solo un 18,59% lo hacen, lo que podría ser positivo (siempre que la etiqueta también este presente) ya que ofrecería un vocabulario añadido. Y que menos de la mitad de la muestra de etiquetas (41,33%) están presentes en el texto variando su morfema. Tanto la aparición de sinónimos como distintas formas morfológicas de la etiquetas son cuestiones positivas, ya que mejoran la recuperación de información gracias a este vocabulario adicional y facilitan la comprensión del usuario/lector.

En nuestro estudio estas variables se han analizado de forma independiente ya que una etiqueta puede estar simultáneamente en el texto, disponer de sinónimos y variar en su morfología en un mismo trabajo. Sin embargo, en el estudio de Heckner et al., (2008) fueron analizadas conjuntamente en función de si las etiquetas estaban presentes de forma idéntica, no idéntica o variando. Los resultados obtenidos fueron (gráfico 11):

- El 54% de las etiquetas aparecieron de forma idéntica en el recurso. Un resultado alejado del nuestro ya que casi la totalidad de las etiquetas de la muestra (94,08%) están presentes dentro de los TFMs.
- El 30% de las etiquetas o no estaban presentes en el texto o aparecían como sinónimos, hiperónimos o hipónimos. Si en este estudio sumamos las etiquetas que no aparecen en el texto, más las etiquetas con sinónimos, obtenemos un porcentaje del 24,51%. Un valor parecido, aunque es normal que nuestro resultado sea menor ya que Heckner et al., (2008) analiza más características en esta variable.
- El 16% de las etiquetas variaban morfológicamente en el texto. En este grupo también tuvieron en cuenta las etiquetas con errores ortográficos. Si nosotros sumamos las etiquetas con variación morfológica más las que presentan algún error ortográfico, obtenemos un porcentaje del 41,85%.

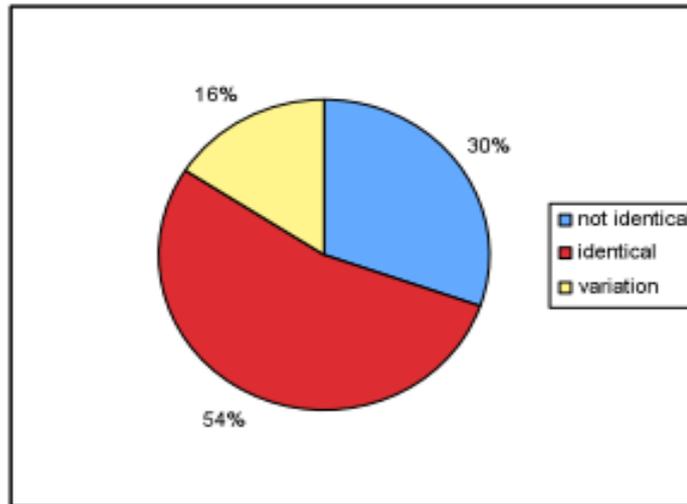


Gráfico 11. Distribución porcentual de las etiquetas en función de su aparición en el texto. Resultados obtenidos por Heckner et al. (2008)

Los resultados en general son bastantes dispares, aunque es cierto que se han utilizado unos criterios metodológicos distintos en ambos trabajos.

Ubicación de la etiqueta dentro del TFM

Se analiza la ubicación y recurrencia de las etiquetas dentro del TFM. Estas pueden aparecer en más de una posición (título, resumen y cuerpo), excepto las etiquetas que solo estén presentes en el apartado de palabras clave (normalmente localizado después del resumen). Las etiquetas que aparezcan en los diferentes apartados con distinto número gramatical o utilicen algún conector (determinantes, preposiciones, etc.) para unir etiquetas compuestas, también han tomado valor en estas variables.

- Etiquetas en “Título”

De las 963 etiquetas que forman la muestra, se ha obtenido que al menos 323 etiquetas aparecen en los títulos de los TFMs, representando un 33,54%.

- Etiquetas en “Resumen”

De las 963 etiquetas que forman la muestra, se ha obtenido que al menos 598 etiquetas aparecen en los resúmenes de los TFMs, representando un 62,10%.

- Etiquetas en “Cuerpo”

De las 963 etiquetas que forman la muestra, se ha obtenido que al menos 854 etiquetas aparecen en el cuerpo de los TFMs, representando un 88,68%.

- Etiquetas solo en “Palabras clave”

De las 963 etiquetas que forman la muestra, se han obtenido 54 etiquetas que aparecen solo como palabras clave de los TFMs, representando un 5,61%.

Apartados del TFM	Ocurrencias	%
Título	323	33,54%
Resumen	598	62,10%
Cuerpo	854	88,68%
Solo en Palabras Clave	54	5,61%
Total muestra	963	

Tabla 6. Total de etiquetas presentes en los diferentes apartados del TFM

Como muestra la tabla 6, las etiquetas que aparecen en el cuerpo de los trabajos alcanzan casi el 90%, un valor positivo ya que si la etiqueta no aparece en el desarrollo del trabajo quizás no esté identificando de forma adecuada el TFM. Los demás resultados también se consideran adecuados, ya que hay que tener en cuenta la extensión reducida tanto del título como del resumen, pues que un 33,54% y 62,10% de etiquetas estén presentes en el título y resumen respectivamente, indican que los autores intentan extraer etiquetas de conceptos que estén presentes en estos apartados tan importantes dentro de los trabajos científicos y de investigación.

Heckner et al., (2008) analizaron también la presencia de etiquetas en estos apartados, aunque utilizaron un criterio diferente al nuestro ya que una vez que la etiqueta estaba presente en unos de los apartados, esta ya no era buscada en los otros restantes. Obtuvieron los siguientes resultados:

- El 49% de las etiquetas aparecen en el título del recurso.
- El 9% de las etiquetas aparecen en el resumen del recurso.
- El 42% de las etiquetas están presentes en el texto del recurso.

Estos resultados difieren a los obtenidos en el presente trabajo, ya que obtienen en su análisis más etiquetas presentes en el título, que en el texto o cuerpo del recurso, además de un valor muy bajo de etiquetas ubicadas en el resumen.

Sin embargo, Chen & Ke (2013), si obtuvieron un resultado similar al nuestro, ya que el 31,87% de las etiquetas estaban presentes en los títulos de artículos de revistas extraídas de Delicious.

Heckner et al., (2008) expusieron que los apartados de título y resumen son apartados relevantes, sobretodo en trabajos científicos, ya que en bases de datos donde no está disponible el texto completo, el título, resumen y palabras clave son la única fuente de información del recurso. Por lo que es recomendable que las etiquetas estén reflejadas en el resumen y algunas de ellas en el título debido a su reducida extensión (independientemente de la aparición obligatoria de las etiquetas en el texto).

En el estudio de Palhares Moreira & Murilo Stempliuć (2006) se analizaron la presencia de las palabras clave de autor en los títulos de artículos científicos, para detectar si estos representan o no adecuadamente el contenido de los trabajos. El resultado mostró que en el 23% de los casos, las palabras clave no aparecían en los títulos. Y se confirmó la hipótesis establecida inicialmente, de que los títulos que cuentan con al menos una de sus palabras clave, identifican el contenido de forma coherente.

Tipo de etiqueta según el contenido

En esta variable se ha analizado si la etiqueta identifica la materia y temática del trabajo o la metodología aplicada en el TFM.

Como se observa en el gráfico 12, las etiquetas relacionadas con la materia del trabajo son más abundantes ya que cuentan con un total de 821 etiquetas de las 963 que forman la muestra, es decir representan un 85,25% del total. Mientras que las etiquetas relacionadas con la metodología aplicada en los TFM recogen un total de 142 etiquetas, obteniendo un porcentaje de 14,75%.

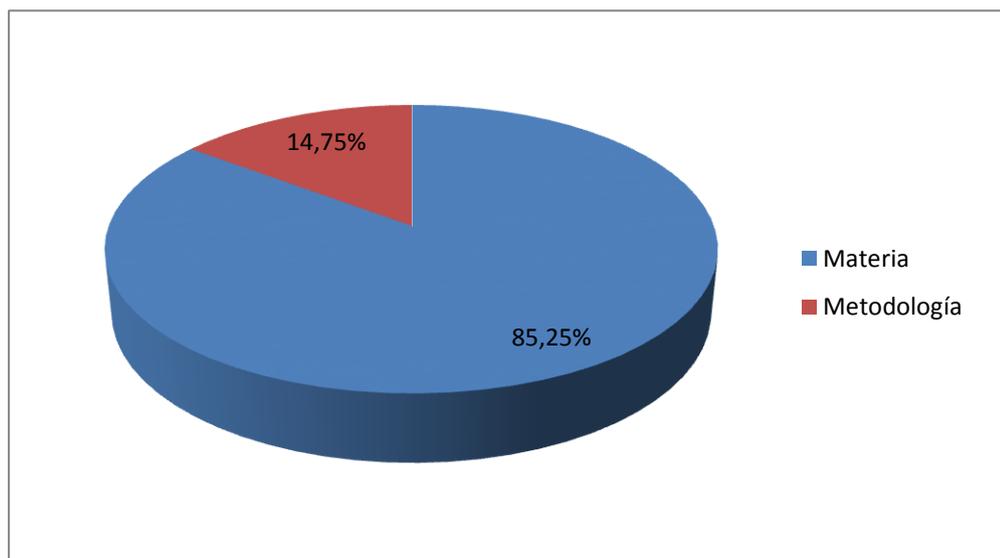


Gráfico 12. Distribución porcentual de las etiquetas en función del contenido que representan

En el estudio Heckner et al. (2008), el 96% de las etiquetas hacen referencia al contenido del recurso y solo un 2% hace referencia a la metodología (el otro 2% restante de la muestra lo clasifica en “códigos” (fechas, número de teléfono, etc.) y “tipo de trabajos” (revisión bibliográfica, tutorial, etc.).

Más recientemente, Choi & Syn (2015) clasifican los recursos etiquetados en materiales textuales, no textuales (imágenes, videos, etc.) y material bibliográfico y observan que las etiquetas relacionadas con el contenido y temática obtienen un 62,89% en textuales, 56,22% en no textuales y 68,04% en materiales bibliográficos. Al enfrentar estos resultados con los nuestros, queda claro que el etiquetador tiende a asignar etiquetas relacionadas con la temática del recurso, pues abundan más que el resto de tipologías.

Lo deseable sería que un recurso sea descrito tanto por etiquetas que hagan referencia a la materia como a la metodología, para representar de una forma completa al recurso y otros usuario al observar las palabras clave sepan de que trata y como se ha realizado el trabajo.

Coincidencia de las etiquetas con las presentes en el Repositorio GREDOS

De las 963 etiquetas analizadas, 621 (64,49%) son respetadas en GREDOS y 342 etiquetas (35,51%) son modificadas o eliminadas.

La coincidencia debería ser de un 100%, ya que en un principio GREDOS respeta las palabras clave de autor. El problema surge cuando varias personas (técnicos de repositorio y bibliotecarios de diferentes facultades) incorporan TFM en GREDOS, por lo que algunos de estos profesionales no creen adecuadas estas palabras clave y las intentan “corregir” y normalizar asignando materias que estén admitidas y signifiquen los mismo. Este tipo de actuación hace que la esencia del etiquetado en lenguaje natural se pierda.

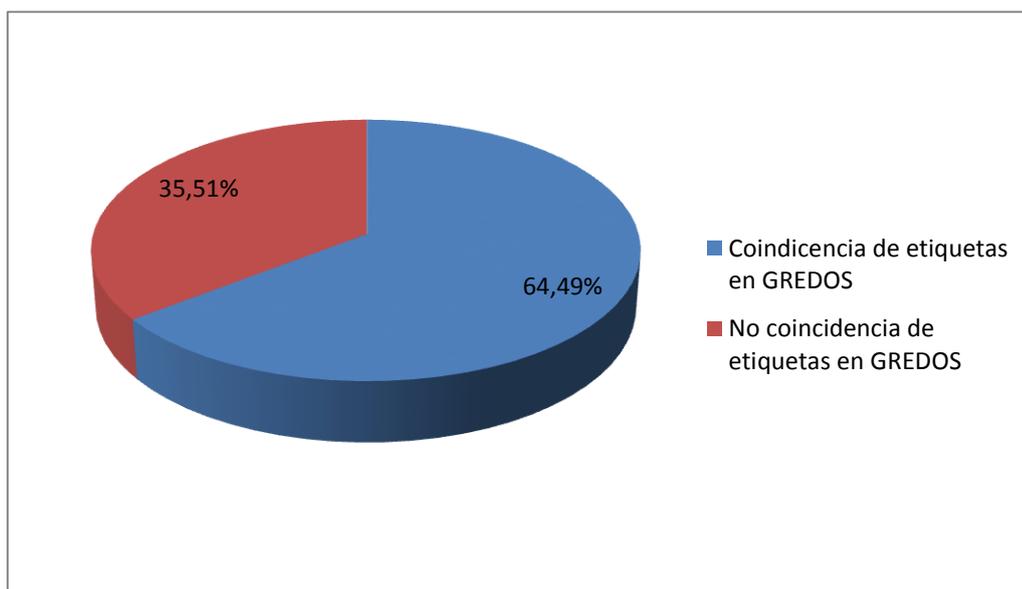


Gráfico 13. Distribución porcentual de la coincidencia y no coincidencia de las etiquetas en GREDOS

Un estudio donde se analizó una cuestión parecida fue el de Gil-Leiva & Alonso-Arroyo (2005), que tras la comparación de palabras clave de autor en sus artículos científicos, frente a los descriptores asignados a estos mismos en varias bases de datos, se alcanzaron los siguientes resultados:

- En IME, el 36,43% de las palabras clave de autor coincidieron con los descriptores.
- En ISOC, el 31,07% de las palabras clave de autor coincidieron con los descriptores.
- En ICYT, el 27,42% de las palabras clave de autor coincidieron con los descriptores.

Otro estudio donde se analizó esta cuestión fue el realizado por Tomás-Casterá et al., (2009). Compararon las palabras clave de autor utilizadas en los artículos publicados en las revistas de nutrición indizadas en la Red SciELO, con los descriptores de Ciencias de la Salud (DeSC), y obtuvieron que de las 3093 palabras clave de autor recopiladas, solo coincidían 988 (31,94%) con descriptores.

Estos resultados obtienen unos valores más bajos que el obtenido en el presente trabajo, ya que GREDOS respeta (no modifica ni elimina) más de la mitad de las etiquetas analizadas (64,49%), aun así este resultado no significa que las palabras clave sean correctas y que los bibliotecarios no crean que deberían ser modificadas.

Si tenemos en cuenta los resultados en general, se llega a la conclusión de que las palabras clave/etiquetas asignadas en lenguaje natural aun no alcanzan el nivel de descripción que proporcionan los vocabularios controlados, aunque estos están limitados por su falta de actualización terminológica.

Número de etiquetas por TFM

En esta variable se ha analizado el número de etiquetas que los autores han añadido a cada TFM, para ver el cual es la cantidad más utilizada en la descripción de TFM. También se ha calculado una media de uso.

Número de etiquetas por TFM	Total TFMs	%
2	2	1,07%
3	24	12,83%
4	44	23,53%
5	56	29,95%
6	36	19,25%
7	14	7,49%
8	1	0,53%
9	5	2,67%
10	3	1,60%
14	1	0,53%
20	1	0,53%
Total TFMs analizados	187	100,00%

Tabla 7. Número de etiquetas por TFM

En la anterior tabla se observa como en 56 TFMs, se han utilizado 5 etiquetas de descripción alcanzando el porcentaje más alto de aparición con un 29,95%, seguido de 44 TFMs con 4 etiquetas y un 23,53% del total de TFMs analizados. Utilizar 4 y 5 etiquetas es lo más común en este contexto, aunque también existen autores que han asignado a sus trabajos, 10, 14 y 20 etiquetas. No por añadir más palabras clave el recurso va a estar mejor descrito, ya que puede que algunas de las etiquetas aplicadas sean redundantes o no sean relevantes respecto al contenido.

Al realizar la media del total de etiquetas recopiladas entre el total de TFMs analizados, se constata un valor de 5,1 etiquetas por TFM. Gil-Leiva & Alonso-Arroyo (2005) observaron que los autores proporcionaban entre tres y cinco palabras clave en los artículos de Ciencias Sociales y Humanidades y Ciencia y Tecnología y entre dos y cinco en el área médica. Heckner et al. (2008), alcanzaron un promedio de 6 palabras clave por artículo en Connotea y Rodríguez Roche & Pérez Sanchidrián (2014), en su análisis de blogs, obtuvieron un margen de entre 4 y 6 etiquetas por post.

Por lo tanto, poniendo en común estos resultados, los etiquetadores tienden a utilizar entre 4 y 6 palabras clave por recurso. Este resultado es una cifra orientativa, ya que lo que realmente importante en la creación y asignación de etiquetas no es la cantidad, sino la calidad de las mismas y su capacidad de descripción.

De forma general, los resultados de este trabajo y los de los estudios que han sido comparados son bastantes parecidos aunque existen diferencias originadas principalmente por la metodología llevada a cabo. Aun así, a pesar de la diferencia en cuanto a tipología documental existente en el conjunto de estudios utilizados, es evidente la existencia de un patrón de etiquetado llevado a cabo por los usuarios/autores.

5 Conclusiones

Acorde a la finalidad y objetivos de este Trabajo Fin de Máster, donde se han analizado las palabras clave de autor de la colección de TFMs ubicados en GREDOS, se enumeran las conclusiones más destacadas:

- La primera conclusión a la que se llegó tras observar el alto número de TFMs sin palabras clave de autor, es a la necesidad de establecer un reglamento único por la USAL, donde se implante un modelo de presentación de TFMs válido para todos los másters, que obligue la asignación de palabras clave, ya que son un elemento básico para la visibilidad y recuperación de todo trabajo científico.
- Respecto a la forma y características gramaticales de la etiquetas se destaca que los autores no tienen una preferencia clara hacia la utilización de etiquetas simples y compuestas en la descripción de TFMs, pues se han obtenido resultados muy similares. Pero si hay que establecer un patrón de comportamiento, los autores utilizan (por muy poca diferencia) etiquetas simples, seguido de etiquetas compuestas formadas por dos términos, y además utilizan una media de 5 palabras clave por TFM. Dentro del análisis gramatical de las etiquetas simples, los sustantivos (nombres común y nombre propio) forman más de la mitad de la muestra. Este valor junto al uso mínimo de adjetivos y verbos es interpretado como positivo, ya que los sustantivos son los que mayor capacidad de representación tienen, y los adjetivos y verbos tienden a ser utilizados en etiquetados egoístas y personales. Sin embargo, el uso de acrónimos sí presenta un problema en la muestra analizada, ya que un mismo acrónimo es asignado de diferentes formas, por lo que esta cuestión debería ser controlada.
- En el análisis de las características lingüísticas de las etiquetas, se destaca la baja existencia de etiquetas en inglés y portugués, que además GREDOS soluciona incorporando en sus registros sus traducciones. Igualmente se reafirma la actitud conservadora de los autores ya que también se ha obtenido un valor bajo de etiquetas consideradas como neologismos, y el cuidado de los autores al asignar etiquetas pues la presencia de errores ortográficos ha sido mínima. Y finalmente se recalca la preferencia del autor en utilizar etiquetas en singular, aunque un cuarto de las etiquetas analizadas están en plural. Por lo tanto, de los principales problemas de la indización libre como son los errores ortográficos, las etiquetas en varios idiomas y la inconsistencia en el número gramatical, el único al que se le debe prestar un mayor grado de atención en el repositorio es al último.
- Respecto a la aparición de la etiqueta dentro del TFM se ha observado como casi la totalidad de las palabras clave de autor aparecen de forma exacta en el trabajo. La presencia de sinónimos de las mismas no destaca significativamente, en cambio, casi la mitad de las etiquetas analizadas presentan algún tipo de variación morfológica dentro del trabajo. Tanto la aparición de sinónimos como las distintas formas morfológicas de las etiquetas son cuestiones positivas, ya que ofrecen un vocabulario adicional

que facilita la comprensión del usuario/lector. Sin embargo, existe un número reducido de autores que describen sus trabajos con palabras clave que posteriormente no aparecen durante el trabajo. En estos casos, es probable que no se esté etiquetando el contenido de una forma correcta, al menos desde un punto de vista de la eficacia y especificidad de la indización.

- Tras analizar la ubicación de las etiquetas, se ha observado como a pesar de la extensión reducida del título y del resumen, los valores de aparición de palabras clave en estos apartados han sido considerados como adecuados. Además, como era de esperar, casi la totalidad de las etiquetas están presentes en el cuerpo de los TFM. Esto se puede deber que el autor, después de crear el trabajo, tiene en cuenta el título y resumen como fuente de extracción de palabras claves, ya que estos apartados son muy importantes dentro los trabajos científicos y de investigación.
- En relación con el contenido que representa la etiqueta, se ha comprobado que el autor tiende a asignar etiquetas relacionadas con la materia y temática del recurso, pues supera con una gran diferencia a las etiquetas metodológicas. Lo correcto sería que un recurso sea descrito tanto por etiquetas que hagan referencia a la materia como a la metodología, para representar de una forma completa el TFM.
- Al analizar la intervención de GREDOS sobre las palabras clave de autor en sus registros, se ha obtenido que en un 35% de los casos, las palabras clave de autor del TFM y las presentes en su correspondiente registro en GREDOS no coinciden. Por lo tanto, este resultado confirma que técnicos de repositorio y bibliotecarios de diferentes facultades, en algunos casos eliminan o modifican las palabras clave de autor, para intentar “arreglarlas” y normalizarlas asignando materias admitidas y que signifiquen los mismo. En consecuencia, con esta intervención profesional la esencia del etiquetado en lenguaje natural va desapareciendo.

Como recomendación, GREDOS debería establecer unas claras instrucciones al respecto. Y como únicas tareas que puede realizar sobre las palabras clave de autor serían las siguientes:

- Incorporación de traducciones al castellano de palabras clave cuando no sean añadidas por el autor.
- Supervisión de las palabras clave para corregir posibles errores ortográficos.

Por lo tanto, si reflexionamos ante estos resultados se llega a la conclusión de que las palabras clave/etiquetas asignadas en lenguaje natural, aun presentan carencias que deben resolverse, ya que por el momento no alcanzan el nivel de descripción que proporcionan los vocabularios controlados, aunque estos también cuentan con sus respectivas limitaciones, como por ejemplo su falta de actualización y cobertura terminológica.

Tanto la descripción en lenguaje natural como la controlada cuentan con ventajas e inconvenientes que podrían aprovecharse y complementarse conjuntamente, ya que los principales problemas de la indización libre (sinonimia, polisemia, singular/plural, etc.) paradójicamente se solucionan con vocabularios controlados, y la indización controlada necesita flexibilidad y adaptabilidad que son la clave de éxito de las folksonomías, es decir, se propone establecer una aproximación de ambas técnicas.

Aun así, independientemente de establecer un vínculo entre la indización libre y controlada, es una realidad que el etiquetado libre y las folksonomías van incrementando y expandiendo su protagonismo en entornos digitales, donde el crecimiento de la información es imparable, y las etiquetas/palabras clave son una herramienta básica de unión entre contenidos y usuarios. Por lo tanto, para que esta tarea obtenga mejores resultados y se minimicen sus errores, se recomienda educar y alfabetizar a los usuarios/autores/etiquetadores dentro de este entorno.

Por consiguiente, se proponen las siguientes recomendaciones a partir de los resultados obtenidos, dirigidos a futuros autores de TFMs de la USAL:

- Utilizar preferiblemente etiquetas simples o etiquetas compuestas por un número reducido de términos
- Evitar utilizar guiones en etiquetas compuestas u otros sistemas de separación diferente al espacio.
- Al utilizar acrónimos como etiquetas, añadir sus siglas desarrolladas posteriormente entre paréntesis, para usuarios que no tengan conocimiento sobre la materia. Ejemplo: *EVA (Entornos Virtuales de Aprendizaje)*
- Asignar acrónimos siempre en mayúscula.
- Asignar etiquetas en castellano y en otro idioma (inglés o portugués). Si el TFM está desarrollado en otro idioma diferente al castellano, asignar etiquetas en ese idioma.
- Utilizar preferiblemente etiquetas en singular.
- Las etiquetas deben estar presentes al menos en los apartados de resumen y cuerpo, y si es posible añadir una (o más) etiqueta en el título del trabajo.
- Asignar etiquetas que hagan referencia a la materia y a la metodología del trabajo.
- Utilizar entre 4 y 6 etiquetas por TFM.

En conclusión, actualmente la indización llevaba a cabo por profesionales a partir de vocabularios controlados no debe ser reemplazada por la indización realizada por autores en lenguaje natural.

Finalmente, para futuras líneas de trabajos, se propone estudiar la posibilidad de crear vías o entornos donde coexistan ambas técnicas, para corregir sus desventajas e incrementar sus beneficios y conseguir una descripción, gestión, distribución y recuperación de información adecuada. También sería interesante realizar estudios que analicen el comportamiento real de búsquedas por parte de los usuarios, así como encuestas de satisfacción de los mismos para conocer si a partir del uso de etiquetas y palabras clave el usuario obtiene resultados relevantes a su consulta, o si por el contrario, al utilizar el lenguaje natural el ruido documental aumenta, es decir, si las etiquetas reducen o incrementan la eficiencia en la búsqueda. El objetivo sería conocer si los usuarios están conformes con los resultados obtenidos en la recuperación. De esta forma el etiquetado libre sería analizado desde diferentes perspectivas, en este trabajo se ha analizado como el usuario asigna etiquetas, y en futuros estudios podría analizarse como posteriormente utiliza estas mismas en el proceso de recuperación de información

Bibliografía

Alonso Arévalo, J. (2009). Los Gestores de Referencias Sociales: índices de popularidad y descubrimiento científico. Recuperado el 17 de abril de 2017 a partir de: <https://gredos.usal.es/jspui/handle/10366/83071>

Alonso Arévalo, J., Cordón García, J. A., & Martín Rodero, H. (2010). CiteULike y Connotea herramientas 2.0 para el descubrimiento de la información científica. *El profesional de la información*, 19(1), 86-94. Recuperado el 17 de abril de 2017 a partir de: <http://eprints.rclis.org/15167/1/EPIGRS.pdf>

Alonso Arévalo, J., Cordón García, J. A., & Martín Rodero, H. (2012). Investigación 2.0 con gestores de referencias sociales: Mendeley y Biowizard. *CIBERMEDICINA*, 2(2), 30-34. Recuperado el 17 de abril de 2017 a partir de: <https://gredos.usal.es/jspui/handle/10366/118456>

Bogers, T., & Petras, V. (2015). Tagging vs. Controlled Vocabulary : Which is More Helpful for Book Search ? En *iConference 2015 Proceedings*. Recuperado el 20 de marzo a partir de: https://www.ideals.illinois.edu/bitstream/handle/2142/73673/65_ready.pdf?sequence=2

Boixados Sanuy, J. (2012). Evaluating the usability of a tag-based, multi-faceted knowledge organization system. Universitat Politècnica de Catalunya. Recuperado el 25 de febrero a partir de: <http://upcommons.upc.edu/handle/2099.1/16388>

Cabré, M. T. (1995). La terminología hoy: concepciones, tendencias y aplicaciones. *Ciência da Informação*, 24(3). Recuperado el 27 de febrero a partir de: <http://revista.ibict.br/ciinf/article/view/567>

Cañadas, J. (2006). Tipologías y estilos en el etiquetado social. En: Yedid, N. (2013). Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. *Información, cultura y sociedad*, (29), 13-26. Recuperado el 15 de febrero a partir de: <http://www.scielo.org.ar/pdf/ics/n29/n29a02.pdf>

Chen, Y. N., & Ke, H. R. (2013). An analysis of users ' behaviour patterns in the organisation of information: A case study of CiteULike. *Online Information Review*, 37(4), 638-656. Recuperado el 25 de abril a partir de: <http://doi.org/10.1108/OIR-05-2012-0086>

Chi-Shiou, L., & Yin Fan, C. (2012). Examinando el comportamiento del etiquetado social y la construcción de una línea de la folksonomía, perspectivas del capital cultural y el capital social. *Ciencias de la Información*, 38(6), 540-557. Recuperado el 23 de marzo a partir de: <https://es.scribd.com/document/327031337/57-Examinar-El-Comportamiento-Etiquetado-Social-y-La-Construccion-de-Una-Linea-Folksonomia-Desde-Las-Perspectivas-de-Capital-Cultural-y-Capital-Social>

- Choi, Y., & Syn, S. Y. (2015). Characteristics of Tagging Behavior in Digitized Humanities Online Collections. *Journal of the Association for Information Science and Technology*, 67(5), 1089-1104. Recuperado el 25 de marzo a partir de: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23472/full>
- Cornella, A. (1999). Cómo sobrevivir a la infoxicación. En *Trascripción de la conferencia del acto de entrega de títulos de los programas de Formación de Posgrado del año académico 1999-2000*. Recuperado a partir de http://www.infonomia.com/img/pdf/sobrevivir_infoxicacion.pdf
- Danielle, L., & Schleyer, T. (2012). Social Tagging Is No Substitute for Controlled Indexing : A Comparison of Medical Subject Headings and CiteULike Tags Assigned to 231, 388 Papers. *Journal of the American Society for Information Science and Technology*, 63(9), 1747-1757. Recuperado el 25 de marzo a partir de: <http://onlinelibrary.wiley.com/doi/10.1002/asi.22653/full>
- Díaz Piraquive, F. N., Joyanes Aguilar, L., & Medina García, V. H. (2009). Taxonomía, ontología y folksonomía, ¿qué son y qué beneficios u oportunidades presentan para los usuarios de la web?. *Universidad & Empresa*, 11(16), 242-261. Recuperado el 10 de enero a partir de: http://www.urosario.edu.co/urosario_files/28/286d2760-7665-49d3-825b-5e43325fb5d4.pdf
- Doerfel, S., Zoller, D., Singer, P., Niebler, T., Hotho, A., & Strohmaier, M. (2016). What Users Actually Do in a Social Tagging System: A Study of User behavior in BibSonomy. *ACM*, 10(2), 32. Recuperado el 20 de marzo a partir de: http://markusstrohmaier.info/documents/2016_ACMTWEB_Bibsonomy.pdf
- Ferreras Fernández, T., & Casaus Peña, A. (2009). GREDOS: Gestión del Repositorio Documental de la Universidad de Salamanca. Recuperado el 23 de marzo a partir de: <http://hdl.handle.net/10366/74739>
- Gil-Leiva, I. (2003). Sistema para la Indización Semiautomática (SISA) de Artículos de Revista de Biblioteconomía y Documentación. En: *II Jornadas de Tratamiento y Recuperación de Información* (pp. 228-232). Recuperado el 18 de enero a partir de: <http://webs.um.es/isgil/resources/SISA%20Indizacion%20automatica%20Automatic%20Indexing%20GI%20LEIVA.pdf>
- Gil-Leiva, I., & Alonso-Arroyo, A. (2005). La relación entre las palabras clave aportadas por los autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. *Revista Española de Documentación Científica*, 28(1), 62-79. Recuperado el 10 de enero a partir de: <http://doi.org/10.3989/redc.2005.v28.i1.165>
- Golder, S., & Huberman, B. A. (2006). The Structure of Collaborative Tagging Systems. Recuperado el 25 de febrero a partir de: <https://arxiv.org/ftp/cs/papers/0508/0508082.pdf>
- Gómez Díaz, R. (2012). *Etiquetar en la web social*. Barcelona: OUC.

Gredos. (2009). Presentación del Sistema de Gestión del Repositorio Documental de la Universidad de Salamanca (GREDOS). Recuperado el 1 de marzo de 2017 a partir de <http://diarium.usal.es/gredos/2009/06/26/presentacion-del-sistema-de-gestion-del-repositorio-documental-de-la-universidad-de-salamanca-gredos/>

Guy, M., & Tonkin, E. (2006). Folksonomies. Tidying up Tags? *D-Lib Magazine*, 12(1). Recuperado el 20 de marzo de 2017 a partir de: <http://www.dlib.org/dlib/january06/guy/01guy.html>

Hassan Montero, Y. (2006). Indización Social y Recuperación de Información. Recuperado 3 de febrero de 2017 a partir de: http://www.nosolousabilidad.com/articulos/indizacion_social.htm

Heckner, M., Mühlbacher, S., & Wolff, C. (2008). Tagging tagging. Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information*, 9(2). Recuperado el 14 de diciembre de 2016 a partir de: <https://journals.tdl.org/jodi/index.php/jodi/article/view/246/208>

López García, X., & Aneiros Díaz, R. (2009). *Documentación e poder na sociedade da información: o papel dos arquivos e os observatorios de comunicación*. Santiago de Compostela: Consello da Cultura Galega. Recuperado el 3 de marzo de 2017 a partir de: http://consellodacultura.gal/mediateca/extras/CCG_2009_Documentacion-e-poder-na-sociedade-da-informacion-Actas-das-xornadas.pdf

Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Recuperado 18 de marzo de 2017 a partir de: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>

Mcfedries, P. (2006). Technically Speaking: Folk Wisdom. *IEEE Spectrum*, 43(2), 80-80. Recuperado el 15 de febrero de 2017 a partir de: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1584367>

Mejias, U. A. (2005). Tag Literacy [Blog]. Recuperado 16 de marzo de 2017, a partir de <http://blog.ulisesmejias.com/2005/04/26/tag-literacy/#q3>

Moens, M. F. (2002). *Automatic Indexing and Abstracting of Document Texts*. En Saavedra Saldívar, J. R. (2010). El etiquetado social de objetos digitales en bibliotecas universitarias : prácticas recomendables (Tesis de maestría). Universidad Nacional Autónoma de México. Recuperado el 2 de febrero de 2017 a partir de: http://eprints.rclis.org/16287/1/Saavedra_Saldivar_2010_ELIS.pdf

Moreiro González, J. A. (2007). La representación de los contenidos digitales: de los tesauros automáticos a las folksonomías. *En Actas del VI Workshop CALSI*. Recuperado el 20 de febrero a partir de: https://issuu.com/bib.csinfo/docs/moreiro_2007

Moreno Ortiz, A. (2008). Ontologías para la terminología : por qué, cuándo, cómo. *Tradumàtica* (6). Recuperado el 27 de febrero de 2017 a partir de: <http://www.raco.cat/index.php/Tradumatica/article/view/123681>

- Motejo Ráez, A. (2003). Asignación automática de palabras clave en tiempo real. Recuperado el 15 de enero de 2017 a partir de: https://www.researchgate.net/profile/Arturo_Montejo-Raez2/publication/245536130_Asignacion_automatica_de_palabras_clave_en_tiempo_real/links/00b495385d6c1969ae000000.pdf
- O'Reilly, T. (2005). What Is Web 2.0. What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. Recuperado 5 de febrero de 2017, a partir de <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Palhares Moreira, M., & Murilo Stempliac, S. (2006). Reconocimiento de títulos de artículos no concordantes con el contenido a través de la utilización de palabras clave. *Scire*, 12(1), 225-245. Recuperado el 5 de marzo de 2017 a partir de: <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=31926480&site=ehost-live>
- Pérez Sanchidrián, E., Campos Posada, R., & Campos Posada, G. E. (2014). Etiquetado social: un modelo de representación de la información en la blogosfera. *Biblios*, (56), 20-27. Recuperado el 15 de febrero de 2017 a partir de: <http://biblios.pitt.edu/ojs/index.php/biblios/article/view/187>
- Quispe Solis, J. C. (2013). Una forma de organizar el conocimiento en la red: las folksonomías. *Accesbib: revista de bibliotecología y ciencias de la información*, 2(1), 5-13. Recuperado el 30 de enero a partir de: <http://eprints.rclis.org/21040/1/art1.pdf>
- Rodríguez Roche, S., & Pérez Sanchidrián, E. (2014). Aproximación al estudio de la objetivación de la realidad cubana en la blogosfera nacional a través de los términos de representación de la información. *Anales de Investigación*, 10(10), 20-36. Recuperado el 23 de enero de 2017 a partir de: http://eprints.rclis.org/25185/1/03-Artículo_02.pdf
- Rodríguez Yunta, L. (2009). Etiquetado libre frente a lenguajes documentales. Aportaciones en el ámbito de Biblioteconomía y Documentación. En *IX Congreso Iberoamericano de España* (Vol. 2, pp. 832-845). Recuperado a partir de <http://digital.csic.es/handle/10261/12295?mode=full&locale=es>
- Saavedra Saldívar, J. R. (2010). El etiquetado social de objetos digitales en bibliotecas universitarias : prácticas recomendables (Tesis de maestría). Universidad Nacional Autónoma de México. Recuperado el 2 de febrero de 2017 a partir de: http://eprints.rclis.org/16287/1/Saavedra_Saldivar_2010_ELIS.pdf
- Sánchez-Zamora, F., Llamas Nistal, M., & Fernández Iglesias, M. (2010). Propuesta gráfica para la caracterización de la dinámica de etiquetado. En *J. Sánchez (Ed.), Congreso Iberoamericano de Informática Educativa* (pp. 579-586). Santiago de Chile. Recuperado el 6 de marzo de 2017 a partir de: <http://www.tise.cl/volumen6/TISE2010/Documento82.pdf>

Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., ... Riedl, J. (2006). Tagging, communities, vocabulary, evolution. En *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. ACM Press. Recuperado el 2 de marzo de 2017 a partir de: <http://files.grouplens.org/papers/sen-cscw2006.pdf>

Soler Monreal, C., & Gil Leiva, I. (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento : folksonomías, taxonomías y ontologías. *Revista Interamericana de Bibliotecología*, 33(2), 361-377. Recuperado el 3 de febrero de 2017 a partir de: <http://www.scielo.org.co/pdf/rib/v33n2/v33n2a4.pdf>

Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. Recuperado 1 de mayo de 2017 a partir de: <http://www.webology.org/2007/v4n2/a41.html#3>

Tomás-Casterá, V., Sanz-Valero, J., Wanden-Berghe, C., & Landaeta, M. (2009). Descriptores versus Palabras Clave sobre nutrición: aportación a la correcta indización. *Anales Venezolanos de Nutricion*, 22(2), 90-94. Recuperado el 27 de marzo de 2017 a partir de: <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=48499477&site=ehost-live%5Cnhttp://content.ebscohost.com/ContentServer.asp?T=P&P=AN&K=48499477&S=R&D=a9h&EbscoContent=dGJyMNHr7ESep7Q4y9fwOLCmr0yep7ZSr6i4TbWWxWXS&ContentCustomer=dGJyMPGqtUqzqrF>

Vander Wal, T. (2005). Explaining and Showing Broad and Narrow Folksonomies. Recuperado 21 de febrero de 2017 a partir de: <http://www.vanderwal.net/random/entrysel.php?blog=1635>

Vander Wal, T. (2007). Folksonomy Coinage and Definition. Recuperado 21 de febrero de 2017 a partir de: <http://vanderwal.net/folksonomy.html>

Yedid, N. (2013). Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. *Información, cultura y sociedad*, (29), 13-26. Recuperado el 15 de febrero a partir de: <http://www.scielo.org.ar/pdf/ics/n29/n29a02.pdf>

Anexo

ANEXO I. Recopilación de palabras clave de autor analizadas.

Ridge	Árbol ternario
Lasso	Empleadas hogar
Elastic net	Calidad de agua
Sparse	Biplot
Reformas Estructurales México	STATIS
Análisis Multivariante	Uso del tiempo
Análisis de Correspondencias Simple	ENUT Colombia 2012-2013
CHAID	Análisis Estadístico Multivariante
STATIS DUAL	HJ Biplot
Análisis Parcial Triádico	Análisis Triádico
Tipos de Tiempo	Inteligencias múltiples
Desarrollo Sostenible	Perfiles profesionales
Responsabilidad Social Corporativa	Instrumento TIMI
Global Reporting Initiative	Big data
Biplot Logístico Externo	Estadística Multivariante
Brasil	Geografía cuantitativa
MMSE	GISciencia
DIF	STATIS
Regresión logística binaria	HJ-Biplot
Analisis de Cluster	Precariedad
Bietápico	Sociología
Two-Step	MDS
HJ-BIPLLOT	Unfolding
ClusPlot	Modelos de Preferencia
Desarrollo sustentable	Soluciones Triviales y Degeneradas
Biplot logístico Externo	Unfolding de Tres Vías
GRI	HJ-Biplot
Empresas mexicanas	Bootstrap
Indicadores	Intervalos de confianza
STATIS	Percentiles
Rendimiento académico	Resinas dentales
Modelo pedagógico	Espectrofotometría
Desempeño académico	Coordenadas CIELab
Dimensiones latentes	Color dental
Biplot	CIEDE2000
Métodos segmentación	CHAEA
Chaid	Biplot
Taid	Análisis factorial confirmatorio

TRI	Metodología cuantitativa
Medicago sativa L.	Asociacionismo en salud
Dactylis glomerata L.	Ayuda Mutua
MANOVA-Biplot	Etnografía
RDA	Artritis Reumatoide
Metaanálisis	Comunidades afrocolombianas
Segmentación	Desplazamiento forzoso
Biología térmica	Reparación integral
Termorregulación	Medicina tradicional
Lacertidae	Mecanismos propios
Literatura Médica	Ley de Víctimas
Medicina Basada en la Evidencia	Cárcel de mujeres
Estadística	Cultura interna
Revisiones Sistemáticas	Prisonalização
Biplot	Reincidencia delictiva
HJ Biplot	Interculturalidad
AMMI	Educación superior
GGE	Educación a distancia
GEI	Vale do Amanhecer
Análisis Factorial	Religión
Impacto	Género
TRI	Prendas
CVRS	Ceará
QUALLEFFO	Brasil
Biplot	Sombrerería
Turismo	Sombrero
Jalisco	Sombrero
Políticas públicas	Cercana desaparición del sombrero
Desarrollo económico	Antiguo oficio de moda
Instituciones	Antropología de la salud
Centralismo	Cirugías plásticas
Apertura económica	Tratamientos postoperatorios
Regiones de Frontera	Belleza
Balance Neto	Salud
Regulación	Epidemiología Sociocultural
Energía eléctrica	Metodología cualitativa en salud
Eficiencia energética	Ritual
CNE	Indígena Zoró
Tarifas eléctricas	Jóvenes
Peajes eléctricos	Edad adulta
VPH	Rito de Paso
Salud Sexual y Reproductiva	Valiente
Antropología Médica	Performance

Sistemas Información Geográfica (SIG)	Devenir
Modelos Nicho Ecológico	Sufrimiento existencial
Biodiversidad	Individuo
Conservación	Restitución
Castilla y León	Vida
Anfibios	Vivencia
Flora interés	Conexión
Crimen organizado	Estructura
Colaboración procesal	Totalidad
Investigación	Valor
Proporcionalidad	Sentido y significado
Seguridad	Conciencia Histórica
Corrupción Pública	Enigma de la vida y del mundo
Impunidad	Concepción de la vida y del mundo
Policía	Autonomía
Brasil	Tecnología
España	Producción
Cooperación civil internacional	Aplicación
Recuperación de activos de la corrupción	Regulación
Decomiso sin condena	Raymond Carver
Responsabilidad civil	Cuentario
Enriquecimiento sin causa	What We Talk About When We Talk About Love
Cooperación directa	Fragmentación
Cooperación Internacional	Indeterminación
Cooperación descentralizada	Multiplicidad
Gobiernos Locales	Polifonía
Políticas Públicas	Afroamericano
República Argentina	Gaston Bachelard
Provincia de Buenos Aires	Hegemonía blanca
Provincia de Corrientes	Hogar
Persona jurídica	Infancia
Abuso de personalidad	Racismo
Levantamiento del velo	Toni Morrison
Espacio Europeo de Educación Superior	Violencia
Sistemas Garantía de Calidad	Ética
Agencias de Calidad	Género
Verificación	Violencia
Títulos de Máster	Sexualidad
Clases particulares	Amor
Shadow education	Política Exterior Brasileña
Variables	América del Sur
Estudio comparado	Soft Power
Academias	Hard Power

Populismo	Colombia
Clase social	España
Voto	Orientación
Hugo Chávez	O-mapa
Evo Morales	LiDAR
Rafael Correa	Penetrabilidad
Venezuela	Vegetación
Bolivia	Data assimilation
Ecuador	Ensemble Kalman filter
Zona de frontera	Parameter adjustment
Cooperación transfronteriza	Parameter identification
Europa	Wildland forest fire modelling
América Latina	Wildfire
Actores locales	GIS
Relaciones binacionales	ArcGIS
Conflicto Armado Colombiano	Arcpy
Colonialidad	Fotogrametría
Representaciones	LIDAR
Identidad	PNOA
Interculturalidad	SNCZI
Cine indígena	Calibración de cámaras
Competencia electoral	Sistemas de Información Geográficos
Sistema de partidos	SIG
Partidos políticos	HEC-RAS
Dimensión vertical	HEC-GeoRAS
Dimensión horizontal	Modelos hidráulicos
Costa Rica	Modelos Digitales del Terreno
Nicaragua	MDT
Comunidad virtual	Modelos Digitales de superficie
Astronomía amateur	MDS
Cultura científica	Modelos Digitales de Elevación
Cultura tecnológica	MDE
Etnografía	ArcGIS
Comunal	PhotoScan
Anticomunal	Láminas
Conjunto de Derechos	Inundación
Sobreexplotación	LiDAR
Infrautilización	Clasificación
Participación Ciudadana	Reconstrucción
Marcos Normativos	Ciudad 3D
Iberoamérica	TerraSolid
Paraguay	Adquisición de la lectura
Bolivia	Personas Sordas Adultas

Identificación de palabras	MOOC
Comprensión oral	Educación Superior
Comprensión lectora	Cursos online
Inclusión educativa	Cursos masivos
Actitudes	Aprendizaje en red
Estrategias	Comunidad de aprendizaje
Docentes	Conectivismo
Necesidades educativas especiales	Adquisición de competencias
Discapacidad	Competencias digitales
Metodología cuantitativa	Medios de comunicación
Empresario-inversor	Habilidades sociales
Rentabilidad-riesgo	Herramientas TIC
Información	Periodismo ciudadano
Recursos gerenciales	E-Learning
Activos reales	Tecnologías de la Información y Comunicación (TIC)
Activos financieros	Entornos Personales de Aprendizaje (PLE)
SICAV	Entornos Virtuales de Aprendizaje (EVA)
Competitive Advantage	Educación Universitaria
Corporate Social Responsibility	Aulas Virtuales
Industry	Scrum
Resource Based View Theory	Aprendizaje basado en proyectos
Research and Development	Patrón pedagógico
Social responsibility	Equipo distribuido
Brand value	Entorno virtual
Stakeholder theory	Evaluación formativa y sumativa
Reputation	Evaluación de aprendizajes
Correlation analysis	e-Evaluación
Fixed effect panel data	TIC
Capital intelectual	Educación secundaria
Evidencia empírica de modelos de capital intelectual	Actitud profesorado
MOOC	E- aprendizaje
COMA	Formación Mixta
Educación superior	E- evaluación
Miríada X	Procedimiento de evaluación
Cursos online	Herramientas EvalCOMIX
Cursos masivos	Herramientas LAMS
Comunidad virtuales	Pensamiento científico en Colombia
Software Educativo	La Ilustración en América
TIC (Tecnologías de la Información y Comunicación)	Museo Nacional de Colombia
Educación Especial	Siglos XVIII y XIX
Necesidades Educativas Especiales	Genética clásica

Genética Molecular	Deterioro cognitivo
Inconmensurabilidad	Diagnóstico
Paradigma Científico	Neuropsicología
Representación mental	Pruebas de rastreo cognitivo
Formación de conceptos	Revisión
Similaridad	Enfermedad de Alzheimer
Espacio conceptual	Fluencia verbal fonológica
Modelo geométrico	Fluencia verbal semántica
Análisis factorial	Memoria semántica
Análisis cluster	Depresión
Deterioro cognitivo leve	Epilepsia
Criterios diagnósticos	Funciones cognitivas
Pruebas neuropsicológicas	Núcleo Subtalámico
Enfermedad del Alzheimer	Núcleo de Meynert
Memoria prospectiva	Enfermedad de Parkinson
Actividad concurrente	Núcleo pálido
Frecuencia de la palabra	Estimulación cerebral profunda
Tiempo de respuesta (TR)	Pobreza
Pista	Exclusión social
Olvido intencional	Formación
Electroencefalograma	Identidad social
Desincronización/ Sincronización cortical relacionada con un evento (ERD/ERS)	Exclusión social
Bereitschaftspotential	Pobreza
Enfermedad de Parkinson	Vivienda
Alteraciones emocionales	Proyecto de intervención
Fase inicial	Barrio Azul
Amígdala	Envejecimiento
Corteza Prefrontal	Dependencia
Hipocampo	Institucionalización
Neuropsicología	Medio rural
Trastorno de estrés postraumático	Cultura Hip Hop
Memoria	Ocio saludable
Amnesia	Ocio nocivo
Rehabilitación	Integración
Intervención	Programas de Transferencia Condicionada
Alteraciones cognitivas	Asignación presupuestaria
Esquizofrenia	Desempeño electoral
Función ejecutiva	Argentina
Neuropsicología	Bolivia
Síntomas negativos	Derechos humanos
Revisión	Calidad democrática

Demencia tipo Alzheimer	Bielorrusia
Autoritarismo	Club de lectura
Participación ciudadana	En línea
Sociedad civil	Nube
Exclusión social	Libros electrónicos
Participación política	Dispositivos electronicos
Participación con carácter económico	Dispositivos Digitales
Participación social	Formatos
Capital social	Perfil del lector
Educación	Plataformas de comercialización
Fracaso Escolar	Amazon
PISA	Casa del Libro
Profesores	Google Play
Opinión	Laie
Préstamo de libros electrónicos	Todoebook
Bibliotecas públicas	Libros electrónicos
eBiblio	Estándares
Castilla y León	Metadatos
Biblioteca pública	Patrimonio cinematográfico
Tecnologías de la información y comunicación	Documentos audiovisuales
Redes sociales	Filmotecas
Capacitación	Delitos informáticos
Facebook	Redes sociales online
Twitter	Clasificaciones especializadas
Personal	Preservación digital
Usuario	Libros electrónicos
Biblioteca Pública de Salamanca Casa de las Conchas	Colecciones digitales
Ediciones Universidad de Salamanca	Bibliotecas Universitarias
Proyecto ENCLAVE Editores-BNE	Recuperación de Información
Biblioteca Nacional de España	Folksonomías
Editoriales universitarias	Twitter
Biblioteca Digital Hispánica	Herramientas 2.0
Wikipedia	Tratamiento documental
Enlace	Web Social
Descripción	CMS
Evaluación	Sistema de Gestión de Contenidos
Marketing en bibliotecas	Arquitectura de la Información
Museo Virtual	Sitios web
Cultura digital	Informe Final del Grupo Incubador de Datos Vinculados de Bibliotecas
Difusión	Grupo Incubador de Datos Vinculados de Bibliotecas
Redes sociales	Realidad aumentada

Casos de uso	Gestión Documental
Conjuntos de datos	Gestión de Contenidos Web
Vocabularios de valores y Conjuntos de elementos de metadatos	Electronic Records Management
Datos Bibliográficos	Workflow
Web 2.0	Colaboration
Biblioteca 2.0	Web
Bibliotecas universitarias	Internet
Revistas electrónicas	Cibermetría
Uso de revistas electrónicas	Crawler
Estudios de usuario	PageRank
Metadatos	Bow-Tie
Patrimonio cinematográfico	Bibliotecas digitales
Documentos audiovisuales	Plantilla de evaluación
Europa	Indicadores
Plataformas de comercialización	Páginas web
Libranda	Trabajo colaborativo
Todoebook	Trabajo en grupo
Leer-e	Edición colaborativa
Laie	Herramientas de software
Amazon	Trabajo científico
Barnes & Noble	Información digital
Libros electrónicos	Publicaciones periódicas españolas
Oferta de libros electrónicos	Investigación
Digitalización	Evaluación
Bibliotecas digitales	Control de autoridades
Google Books	Fichero internacional virtual de autoridades
Europeana	Facebook
Catálogos	Privacidad
Interfaces	Redes sociales
Recuperación de la información	Ley de protección de datos
Evaluación	Datos de carácter personal
Evolución	Libro electrónico
Administración electrónica	Web 2.0
Comunidades autónomas	Biblioteca 2.0
Servicios de Información electrónica	Red social
Unión europea	Servicios de información electrónicos
España	Prensa digital
Políticas de información	Periodismo electrónico
ECM	Preservación digital
DMS	Políticas
Softwre Libre	Recursos
Digitalización	Metadatos de preservación

PREMIS	Cáncer
METS	Cannabis terapéutico
OAIS	Dolor
Repositorios institucionales	Tratamiento
Gredos	Niños
Universidad de Salamanca (España)	Oncología pediátrica
Descripción de recursos digitales	Control
ISBD (ER)	Dolor
ISBD	Terapia de juego
Computación en la nube	Índices
Tecnología de la nube	Análisis
Gestión de la información	Producciones escritas
Seguridad de la información	Correlaciones
Servicios en la nube	Lectura
Ocio	Habilidades lingüísticas
Trastorno del espectro autista (TEA)	Adicción
Discapacidad Intelectual	Uso problemático
Participación	Internet
Inclusión	Teléfono móvil
Calidad de Vida	Salud
App	Adolescentes
Emociones	Calidad de vida
LIWC	Adolescentes
Moodies	Discapacidad Intelectual y del desarrollo
Reconocimiento de emociones	Contexto
Calidad de vida	Autodeterminación
Baja visión	COMA
Ceguera	Aprendizaje ubicuo
Nuevas tecnologías	Experiencia de aprendizaje
Tiflotecnología	Diseño centrado en el usuario
Adaptación	Aprendizaje adaptativo
Neurociencia cognitiva	Educación abierta
Trastornos del espectro autista	Wearable
Atención conjunta	Educación
Alteraciones	Creatividad
Estructuras	Internet
Funciones	Pedagogía
Condicionamiento evaluativo	Patrón pedagógico
Condicionamiento	Lenguaje de patrón
Aprendizaje	Bullying
Conciencia	Ciberbullying
Memoria	Formación del profesorado
Emociones	Inclusión

Interculturalidad	Editoriales universitarias
Acogida de refugiados	Descripción
Inteligencia Emocional	Datos Abiertos
TMMS-24	Gestión de datos
HOMALS	Gobierno Abierto
Biplot	Modelo de negocios
STATIS	Conciencia fonológica
Satisfacción estudiantil	Lenguaje oral
Análisis factorial	Memoria de trabajo
Teoría de respuesta al ítem	Control inhibitorio
Impacto	Lectura
Análisis Biplot	Obras audiovisuales
Exclusión social	Derechos de Autor
Pobreza	Digitalización
Análisis discriminante	On line
HOMALS	Límites
HJ Biplot	Prueba
Comportamiento electoral	Electrónica
Regresión logística multinomial RLM	WhatsApp
Análisis de correspondencias múltiples ACM	Teléfono
Análisis de segmentación	Mensajería
Biplot	Justicia 2.0
Twitter	Obtención
Información y Documentación	Aportación
Redes sociales	Validez
Detección temática	Legalidad
Libros electrónicos	Autenticidad
Literatura infantil y juvenil	Vulnerabilidad
Plataformas de préstamo	Proceso
Aplicaciones de lectura	Penal
Aplicaciones de lectura infantil	Diligencias investigación
Aplicaciones para iOS	Prueba electrónica
Aplicaciones para Android	Delitos informáticos
Literatura infantil	Procesos matrimoniales
Repositorios Institucionales	Procedimiento contencioso
Acceso abierto	Procedimiento de mutuo acuerdo
Universidad de Salamanca (España)	Medios de prueba
Datos Abiertos	Protección del menor
Gobierno Abierto	Audiencia al menor
Reutilización de datos	Mediación Penal
Ediciones Universidad de Salamanca	Justicia Restaurativa
Catálogos	Víctima
Exposiciones	Infractor

Mediador/a	Delito Leve
Despido colectivo	Reforma
Grupo de empresas a efectos mercantiles	Código Penal
Grupo de empresas a efectos laborales	Violencia
Elemento adicional	Discapacidad
Impugnación del despido colectivo	Vulnerabilidad
Propiedad Horizontal	Discriminación
Accesibilidad universal	Responsabilidad patrimonial
Discapacidad	Ámbito sanitario
Eliminación de barreras arquitectónicas	Pérdida de oportunidad
Ascensores	Interés del menor
Impugnación de acuerdos	Desprotección
Abogacía	Riesgo
Cuota litis	Desamparo
Publicidad	Tutela administrativa
Secreto profesional	Reinserción familiar
Régimen disciplinario	Partición hereditaria
Resolución de conflictos	Contador-partidor
Obligatoriedad	Operaciones particionales
Mediación familiar	Colación
Ejecutividad de los acuerdos	Herederó
Custodia	Despido disciplinario
Compartida	Improcedente
Alternativa	Indemnización por despido
Interés	Carta de despido
Menor	Régimen especial
Acuerdo	Abogados asalariados
Abogado	Obligación
Arbitraje	Parientes
Árbitro	Alimentos
Puerto Seguro	Cláusula suelo
Sentencia TJUE 6/10/2015	Empresario
Decisión 1250/2016	Préstamo hipotecario
EEUU	Condiciones
Seguridad	Control de transparencia
Protección de datos	Buena fe contractual
Guarda	Virus espía
Custodia	Diligencia
Compartida	Cadena de custodia
Jurisprudencia	Prueba electrónica
Tribunal Supremo	Bicicleta
Interés del menor	Circulación
Falta	Responsabilidad civil

TEA
Trastorno del Espectro Autista
Autismo
Habilidades sociales
Programas psicoeducativos
Inclusión
Autoeficacia
Género
Diferencias
Decisión de carrera
CTIM (Ciencia, Tecnología, Ingeniería, Matemáticas)
TEA
Conductas repetitivas y restringidas
RBS – R
Escala
Adicción a sustancias
Dependencia emocional o afectiva
Adicción al amor
Dependencia sentimental
Deterioro cognitivo
Funciones ejecutivas
Enfermedad de Parkinson
Rehabilitación neuropsicológica
Bulimia Nerviosa
Adicción a internet
Impulsividad
Pérdida de control
Autocontrol
Síndrome de Burnout
Servicios de Urgencias Hospitalarias
Prevalencia
Maslach Burnout Inventory
Symptom Check List SCL-90-Revised
Transmisión intergeneracional
Trauma
Prevención
Repercusiones psicológicas
Síndrome de Sobrecarga
Cuidadores primarios
Estilos de afrontamiento
Enfermos oncológicos