



VNiVERSIDAD  
D SALAMANCA

Análisis formal del etiquetado social:  
Flickr y Delicious: estudio de caso

MARÍA TRUJILLO HUESO

bajo la dirección de

CARMEN CARO CASTRO

UNIVERSIDAD DE SALAMANCA

FACULTAD DE TRADUCCIÓN Y  
DOCUMENTACIÓN

GRADO EN INFORMACIÓN Y DOCUMENTACIÓN

Trabajo de Fin de Grado

Análisis formal del etiquetado social:  
Flickr y Delicious: estudio de caso

María Trujillo Hueso

bajo la dirección de

Carmen Caro Castro

Salamanca, 2013

TRUJILLO HUESO, María

Análisis formal del etiquetado social: Flickr y Delicious : estudio de caso / María Trujillo Hueso; bajo la dirección de Carmen Caro Castro – Salamanca: Universidad de Salamanca, Facultad de Traducción y Documentación, 2013.

Trabajo de Fin de Grado – Grado en Información y Documentación.

1. Etiquetado social. 2. Folksonomías. 3. Flickr. 4. Delicious. I. Caro Castro, Carmen, dir. II. Título

025

## Resumen

Las tecnologías de la información han provocado un gran cambio en la forma de comunicarse, obtener y transmitir información. Gran parte de esta transformación ha afectado al usuario, quien ha pasado de ser un mero receptor pasivo a ser un agente activo, que tiene la posibilidad de crear y gestionar la información en la web. Uno de los productos que ha generado esta nueva situación son las folksonomías. A pesar de su aceptación como herramienta para representar y recuperar información, no carece de desventajas. La más destacada por dificultar la recuperación de información, es la ambigüedad inherente a todos los vocabularios no controlados.

El objetivo fundamental de este trabajo es cuantificar la aparición de problemas lingüísticos (sinónimos, singular/plural, multilingüismo, expresiones compuestas y errores ortotipográficos) en dos muestras de tags de dos portales (Flickr y Delicious) y analizar su semejanza con el fin de verificar si se ven afectados por los mismos problemas y con la misma intensidad. Para tal propósito, ambas muestras son clasificadas y sometidas a un proceso de limpieza y clasificación; y posteriormente a un análisis cualitativo y cuantitativo que permite comprobar la presencia y el grado de aparición de dichos problemas.

Los resultados obtenidos confirman la presencia de los problemas en ambas páginas, en general con un grado de aparición e influencia similares. En casos determinados las desigualdades existentes se explican por medio de la exposición de características concretas de las páginas. También se comprueba cómo los valores obtenidos se corresponden con resultados de trabajos anteriores similares, sugiriendo la existencia de un patrón en la manera de etiquetar de los usuarios. Se plantean una serie de líneas de trabajo futuras con el propósito de ampliar el estudio y definir más claramente la evolución constante de la folksonomía de ambos portales.

**Palabras clave:** etiquetado social, etiquetas, folksonomías, Flickr, Delicious.

## Abstract

Information technologies have caused a major shift in the way people communicate, obtain and transmit information. Much of this transformation has affected the user, who has gone from being a passive recipient to be an active agent that has the ability to create and manage information on the web. One product that has generated this new situation are folksonomies. Despite its acceptance as a tool to represent and retrieve information, not without disadvantages. The most prominent, which difficult the information retrieval, is the ambiguity inherent in all uncontrolled vocabularies.

The main objective of this work is to quantify the emergence of language problems (synonyms, singular/plural, multilingualism, compound expressions and typographical errors/misspellings) in two tags samples from two portals (Flickr and Delicious) and analyze

their similarity to verify if they are affected by the same problems and with the same intensity. For this purpose, both samples are classified and subjected to cleaning and sorting process, and then to a qualitative and quantitative analysis that enables to check up on the presence and degree of occurrence of these problems.

The results obtained confirm the presence of the problems in both pages, generally with an alike degree of appearance and influence. On certain cases, inequalities are explained by the exposition of specific characteristics of the pages. It is also tested out how the values obtained correspond to results of similar previous work, suggesting the existence of a pattern in the way of labeling performed by the users. Future lines of research are suggested in order to extend the study to more clearly define the constant evolution of the folksonomy created on both portals.

**Keywords:** social tagging, tags, folksonomies, Flickr, Delicious.

## Sumario

Justificación del trabajo	8
1. Marco teórico	10
1.1. Web 2.0	10
1.1.1. La Biblioteconomía en la web 2.0	12
1.2. Las tags	13
1.3. Tagging social o colaborativo	14
1.3.1. Tipos de etiquetado	14
1.4. La Folksonomía	15
1.4.1. Características de las folksonomías	16
1.4.2. Clasificación de las folksonomías	17
1.4.3. Ventajas de las folksonomías	19
1.4.4. Inconvenientes de las folksonomías	20
1.4.5. Problemas lingüísticos de las folksonomías	21
1.5. Dos folksonomías actuales: Flickr y Delicious	25
1.5.1. Características de los portales a estudio	25
2. Metodología	30
2.1. Objetivos	30
2.2. Técnica de recogida de datos	30
2.3. Métodos de análisis	35
3. Resultados	37
3.1. Resultados obtenidos frente a resultados anteriores	52
4. Conclusiones y futuras líneas de trabajo	54
Bibliografía	56
Anexo	60

## Índice de figuras

Figura 1. Tag cloud	13
Figura 2. Mapa conceptual de una Folksonomía	17
Figura 3. Folksonomía amplia	18
Figura 4. Folksonomía estrecha	18
Figura 5. Pantalla de resultados de búsqueda de Flickr	23
Figura 6. Interfaz básica de Flickr	25
Figura 7. Interfaz de subida de imágenes de Flickr	27
Figura 8. Interfaz básica de Delicious	28
Figura 9. Interfaz de subida de enlaces de Delicious	29
Figura 10. Tag cloud de Flickr	31
Figura 11. Muestra de tags relacionadas dentro de una tag principal	31
Figura 12. Muestra de tags relacionadas dentro de una tag principal	33
Figura 13. Detalle de la muestra de tags	33
Figura 14. Resultados de la limpieza de Flickr	34
Figura 15. Resultados de la limpieza de Delicious	34
Figura 16. Detalle de la muestra de tags	34
Figura 17. Valor de aparición de sinónimos en	37
Figura 18. Valores de aparición de sinónimos totales, intra e interlingüísticos	38
Figura 19. Etiquetas afectadas por la dicotomía plural/singular	38
Figura 20. Valores de aparición de la dicotomía plural/singular en Flickr	39
Figura 21. Valores de aparición de los idiomas utilizados en Flickr	40
Figura 22. Valores de aparición de los idiomas secundarios en Flickr	40
Figura 23. Valor de aparición de expresiones compuestas en la tags de Flickr	41
Figura 24. Valor de aparición de sinónimos en Delicious	42
Figura 25. Valores de aparición de sinónimos totales, intra e interlingüísticos	42
Figura 26. Etiquetas afectadas por la dicotomía plural/singular	43
Figura 27. Valores de aparición de la dicotomía plural/singular en Delicious	44
Figura 28. Valor de aparición de errores ortotipográficos en Delicious	44
Figura 29. Valores de aparición de los idiomas más utilizados en Delicious	45
Figura 30. Valores de aparición de los idiomas secundarios más utilizados en Del.	45
Figura 31. Valor de aparición de expresiones compuestas en la tags de Del.	46
Figura 32. Valores de aparición de sinónimos en Flickr y Delicious	46

---

Figura 33. Valores de aparición de sinónimos intra e interlingüísticos	47
Figura 34. Valores de aparición de sinónimos intra e interlingüísticos	47
Figura 35. Porcentaje de etiquetas afectadas por de la dicotomía plural/singular en Flickr y Delicious	48
Figura 36. Valores de aparición de la dicotomía plural/singular en Flickr y Del.	48
Figura 37. Porcentaje de uso del idioma inglés en Flickr y Delicious	49
Figura 38. Valor de aparición de los 5 idiomas principales	50
Figura 39. Valores de aparición de expresiones compuestas en Flickr y Del.	51

### **Índice de tablas**

Tabla 1. Diferencias entre la Web 1.0 y la Web 2.0	11
Tabla 2. Valores totales obtenidos	60



## Justificación del trabajo

La irrupción de Internet y su rápida evolución ha trastocado la manera en que la humanidad se comunica. Sin embargo, para los nativos digitales, Internet y la World Wide Web siempre han estado ahí. Se han ido adaptando a sus cambios e innovaciones con naturalidad. Conceptos como web 2.0, wikis, blogs, comunidades virtuales... resultan tan cotidianos que son aceptados sin plantearse cuestiones relativas al porqué, a su estructura ni al proceso de evolución seguido hasta llegar a este estado.

Con el paso de la web 1.0 a la actual 2.0 el rol del usuario ha sufrido un cambio significativo. Ya no desea ser un espectador pasivo, no se conforma con hacer click en un recurso y visualizarlo. Quiere tener herramientas, más o menos complejas, que le permitan una cierta capacidad de decisión y de control. Es decir, el usuario se convierte en su propio gestor de información.

Una de las principales herramientas de gestión de información utilizada actualmente es el etiquetado social, el cual permite crear contenido web, organizarlo y clasificarlo en categorías por medio de etiquetas (tags). En este sistema de organización son los propios usuarios quienes eligen las tags, vocablos que funcionan como palabras clave y describen el contenido del recurso a compartir.

Este trabajo estudia las folksonomías, que son el resultado de la gestión de información realizada por los usuarios (etiquetado social). Se trata de un vocabulario colectivo y un lenguaje no controlado formado por las tags asignadas libremente por los usuarios. La folksonomía adquiere relevancia por poseer cualidades como espontaneidad, libertad, no regularización..., en oposición a la tradicional taxonomía (*clasificación científica*), en la cual un indizador experto asigna los descriptores en función de unas normas y sistemas de organización, para facilitar el acceso a la información.

La extensa y activa utilización del etiquetado social para gestionar la información, y de las folksonomías como instrumento de recuperación de dicha información, sugiere una amplia aceptación de este fenómeno tan reciente. Sin embargo, no faltan autores que consideran y exponen sus debilidades, poniendo en duda su valor y su conveniencia.

A pesar de la variedad de problemas que se mencionan en la bibliografía, es indiscutible que el de mayor peso es la ambigüedad de las folksonomías, inherente a todos los vocabularios no controlados. Cuanto más ambigua sea la terminología empleada, más difícil será la recuperación de la información, que es el fin perseguido al organizar y categorizar los recursos compartidos. Entonces, ¿hasta qué punto esta ambigüedad está presente en las folksonomías creadas por los usuarios? Este trabajo busca la respuesta analizando en profundidad dos muestras de tags de los portales mencionados, para verificar la presencia de estos problemas, a continuación cuantificar su frecuencia de aparición y, finalmente, contrastar la correspondencia de estos valores con los sugeridos por expertos.

La amplitud del campo de estudio posible, todos los portales que recurren el etiquetado social como sistema de organización, debe acotarse necesariamente. En este caso se han seleccionado dos de sus máximos exponentes, considerados como tales por su popularidad y su amplio uso: Flickr (*imágenes y vídeos*), y Delicious (*enlaces*).

Ya que un análisis en profundidad de todas las tags de estos dos portales requiere una enorme inversión de tiempo, esfuerzo y recursos no disponibles, se impone la necesidad de realizar una selección. Para tal propósito se han recogido dos muestras, una de cada portal, a las que se aplica un proceso de limpieza, organización y categorización. Este último paso permite, además, obtener una perspectiva de los hábitos y preferencias de los usuarios a la hora de escogerlas. Se obtiene una ventaja añadida al aplicar estos procesos a dos portales distintos; la de ser posible realizar una comparación entre los resultados.

A continuación se presenta una introducción teórica a este fenómeno, su origen y características, y los trabajos y estudios que comparten el interés por las folksonomías. Posteriormente, en el apartado de metodología, se expone en profundidad las técnicas de recogida y análisis de datos, se cuantifican los resultados presentándolos en forma de gráficos. Finalmente se realiza una comparación entre los valores de ambos portales antes de exponer las conclusiones.

## 1. Marco teórico

### 1.1. Web 2.0

Son abundantes las referencias a la Web 2.0 en la bibliografía como sinónimo de modernidad, de dinamismo y evolución; cualidades comunes a las denominadas “nuevas tecnologías de la información y la comunicación” abreviadas como “nuevas tecnologías” y el acrónimo TIC. Pero ¿Porqué se denomina 2.0? acaso ¿Existe la 1.0? y, entonces, ¿cuándo veremos la 3.0?

La idea de Web 2.0, el modelo predominante en la actualidad, fue presentada en sociedad en el año 2004 por Tim O’Reilly<sup>1</sup> para referirse al nuevo modelo de Web que a su juicio se imponía, desplazando a su predecesora la Web 1.0. La asignación de una numeración ascendente en la denominación sugiere la idea de dinamismo, de evolución en positivo, de superación del modelo anterior, así como la esperanza de que exista la 3.0 (o Web semántica) y siguientes, en la convicción de que estamos asistiendo a la génesis de un fenómeno tecnológico que está cambiando la forma en que los seres humanos se comunican y relacionan.

El nuevo término fue asumido rápidamente recogándose en la wikipedia<sup>2</sup> como una forma de hacer sitios y aplicaciones web que tienen en cuenta al usuario como creador y receptor de información, interactuando con el sistema y entre los usuarios. Mengual (2009, 4) amplía la definición señalando que se orienta a la *“interacción y redes sociales, que pueden generar contenido que explota los efectos de las redes con o sin crear web interactivas y visuales.....que actúan más como punto de encuentro o web dependiente de usuarios que como web tradicionales”*. Se refiere así a la contraposición entre la Web tradicional (1.0), estática, dirigida a un usuario consumidor no participativo; y lo que es la señal de identidad de la Web 2.0, la participación del usuario. En línea similar Adell Ortega (2009, 4) la identifica como una *“segunda generación de web basadas en comunidades de usuarios y una gama especial de servicios, como las redes sociales, los blogs, las wikis o las folksonomías, que fomentan la colaboración y el intercambio ágil de información entre los usuarios”*. Manifiesta así una mirada integradora y utilitarista de la Web actual, en la que actúa como un catalizador que hace posible el aprovechamiento de todo lo nuevo procedente de corrientes, tendencias, tecnologías y circunstancias. Más lejos llega Serrano (2006) al afirmar que en la Web 2.0 el poder lo tiene el usuario y que su filosofía subyacente es que la plataforma pone el sitio cediendo al usuario la tarea de aportar el contenido.

<sup>1</sup> Es fundador y presidente de la editorial O'Reilly Media, un fuerte impulsor de los movimientos de software libre y código abierto, así como uno de los autores del concepto Web 2.0 y participante en el desarrollo del lenguaje Perl. [http://es.wikipedia.org/wiki/Tim\\_O%27Reilly](http://es.wikipedia.org/wiki/Tim_O%27Reilly) [Fecha de consulta: 3 de agosto de 2012]

<sup>2</sup> Definición disponible en: [http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0) [Fecha de consulta: 2 de junio de 2012]

La tabla 1, que se muestra a continuación, ilustra la evolución producida desde la creación de la web 1.0 hasta la actual 2.0, teniendo en cuenta la información manejada, las características de los sitios web (sus contenidos, administradores y objetivos), y el tipo de software.

Web 1.0	Web 2.0
Información centralizada	Información descentralizada
Información poco actualizada	Información en continuo cambio
Contenidos y sitios más bien estáticos	Contenidos y sitios flexibles, en constante transformación
Sitios con contenidos de alta y baja calidad administrados por un web máster	Amplia diversidad en contenidos administrados por los usuarios
Diseño y producción a cargo de quienes conocen sobre informática	Diseño y producción sin necesidad de grandes conocimientos sobre informática. Accesibles y prácticos.
Sitios con fines generalmente comerciales	Sitios con fines diversos; en la mayoría de casos, la construcción de comunidades que comparten intereses, prácticas, información, etc.
Software tradicional	Software y aplicaciones "on line" que no requieren de su instalación en el PC para utilizarlos

Tabla 1. Diferencias entre la Web 1.0 y la Web 2.0<sup>3</sup>

Aunque la idea de cambio radical de modelo está ampliamente aceptada, otros autores, en línea con el creador de la Web Tim Bernes Lee<sup>4</sup> (2006), no aceptan la existencia de una Web 2.0. Consideran que la intención colaborativa así como la participación inter pares ya estaba presente en la anterior, y prefieren considerar que es la evolución de las aplicaciones las que han hecho posible la generalización de la participación.

Ejemplos de la Web 2.0 son las comunidades web, los servicios web, las aplicaciones web, los servicios de red social, los servicios de alojamiento de videos, las wikis, blogs, mashups y folksonomías.

<sup>3</sup> Fuente: Redes Sociales (blogspot). *Diferencias entre Web 1.0 y 2.0 (parte 2)*. Se han hecho cambios en la presentación para facilitar la comprensión. Disponible en: <http://4esoconsolacionjtorrenteymtoledo.blogspot.com.es/2010/05/diferencias-entre-web-10-y-20-14.html> [Fecha de consulta: 25 de octubre de 2012]

<sup>4</sup> Desarrolló las ideas fundamentales que estructuran la web. Él y su grupo crearon el lenguaje HTML, el protocolo HTTP y el sistema de localización de objetos en la web URL. Se considera al mismo tiempo el inventor y el protector de la web. [http://es.wikipedia.org/wiki/Tim\\_Berners-Lee](http://es.wikipedia.org/wiki/Tim_Berners-Lee) [Fecha de consulta: 10 de agosto de 2012]

### 1.1.1. La Biblioteconomía en la web 2.0

Las bibliotecas, al igual que otros muchos aspectos de nuestra vida, no han quedado al margen de la influencia de las TIC. Al contrario que en la Web 2.0 donde, como veremos, los usuarios clasifican la información, en las bibliotecas sólo los profesionales pueden crear, clasificar y poner a disposición de los usuarios los materiales. Tienen los mejores sistemas de categorización conocidos, pero la evolución progresiva de la Web ha creado una brecha entre los sistemas de clasificación anteriores y los modernos, haciendo que algunos parezcan pasados de moda (Shirky, 2005).

La diferencia en el volumen de información manejada en una biblioteca y en la Web es enorme. En aquella, los materiales mantienen un ritmo de entrada generalmente planificado, hay profesionales dedicados exclusivamente a procesarlos, y el volumen de información se mantiene dentro de unos parámetros manejables. Por el contrario, la Web, sin restricciones, posee un volumen de información casi inabarcable que crece descontroladamente a cada segundo, con millones de usuarios gestionándola sin pautas establecidas. Por ahora, parece utópico pretender una Web totalmente clasificada y ordenada según unos parámetros estrictos.

Desde hace algún tiempo, los profesionales de la documentación se debaten entre la conveniencia de mantener la tradición del uso de los vocabularios controlados, de fiabilidad contrastada, y la oportunidad de incorporar las ventajas de los lenguajes libres, con el inconveniente de su limitada consistencia. Es una cuestión compleja, ya que en general, y probablemente influidos por su estilo de trabajo, los profesionales consideran que el control y la normalización son irrenunciables. Sin embargo, autores como Moreiro (2006) y Lancaster (2002) consideran que los lenguajes libres tienen algunas ventajas sobre los vocabularios controlados:

- Moreiro (2006) considera que no se precisa inversión para construir lenguajes documentales, son lenguajes colaborativos, ofrecen una enorme riqueza de vocabulario, son fácilmente automatizables al trabajar sobre todo con unitérminos, y se obtienen resultados satisfactorios cuando se combinan con los términos propios de un entorno científico-técnico específico.
- Por su parte Lancaster (2002, p. 188) cree que su uso será predominante en el futuro: *“Parece evidente que el lenguaje natural será la norma en la recuperación de información y que el uso de los vocabularios controlados convencionales disminuirá. Existen numerosas razones para ello, como los elevados costes del proceso intelectual humano, la rápida disminución de los costes de almacenamiento automatizado, el creciente volumen de texto que se encuentra accesible por ordenador (incluyendo el correo electrónico y el texto completo de revistas y periódicos), y la reducción gradual de la dependencia de intermediarios cualificados en la búsqueda online”*.

Otros autores consideran que la situación de los tesauros, los encabezamientos de materias y otras herramientas de clasificación se enfrentan a la disyuntiva de “renovarse o morir”. Proponen, para evitar llegar a un punto en que estas herramientas sean inservibles, realizar

una incorporación de las folksonomías a los tesauros, los catálogos de bibliotecas, etc. La manera de integrar la rigidez de los sistemas clásicos con la libertad y la ambigüedad del lenguaje libre, es un tema que continúa en fase de estudio.

## 1.2. Las tags

La primera aparición de las tags<sup>5</sup> (etiquetas) en la Web fue adelantada por Yahoo en la web Open Directory<sup>6</sup> en 1998, iniciando un fenómeno de gran éxito con un crecimiento sorprendente tanto en intensidad como en extensión.

Una tag es una palabra que, asociada a un recurso, actúa como una marca o señal cuya intención es informar sobre alguna cualidad o atributo de ese recurso y que facilita su selección, clasificación o reconocimiento.

Las tags son creadas por los propios usuarios por decisión libre no sujeta a reglas, como reflejo de una suerte de poder popular (Quintarelli, 2005) e incorporadas en páginas web, blog, wikis. A menudo se presentan agrupadas formando una nube (tags cloud) en la que aparecen resaltadas con letras de mayor tamaño las más populares, es decir, aquellas con mayor frecuencia de utilización.



Figura 1. Tag cloud<sup>7</sup>

<sup>5</sup> En este trabajo se usa el término *tag* en inglés por ser el más extendido, en vez de su equivalente en español *etiquetas*.

<sup>6</sup> Open Directory es un proyecto colaborativo multilingüe, en el que editores voluntarios listan y categorizan enlaces a páginas web. Disponible en : <http://www.dmoz.org/> [Fecha de consulta: 17 de noviembre de 2012]

<sup>7</sup> Fuente: *Wikipedia. Tag cloud*. Disponible en: [http://upload.wikimedia.org/wikipedia/commons/a/a7/Web\\_2.0\\_Map.svg](http://upload.wikimedia.org/wikipedia/commons/a/a7/Web_2.0_Map.svg) [Fecha de consulta: 5 de julio de 2012]

Existe un amplio consenso entre los expertos al considerarlo como un nuevo método de indización no controlado, contrapuesto a los métodos tradicionales (tesauros, taxonomías, clasificaciones, etc.). Voß (2007) considera mejor verlo como una indización popular que está contribuyendo al renacimiento de la indización manual.

### 1.3. Tagging social o colaborativo

La facilidad que ofrece la World Wide Web (WWW) para observar, exponer y participar ha disparado la intervención de los usuarios en la creación y asignación de tags a los recursos. La magnitud del fenómeno ha dado lugar al neologismo “tagging social”<sup>8</sup> para nombrar al sistema colaborativo, compartido por otras muchas aplicaciones en la red (Díaz et al, 2009), que permite generar datos producidos por la participación de miles de usuarios. Se entiende como una actividad desarrollada en un contexto sociotécnico que le da sentido, posibilitando el desarrollo del tagging (Trant, 2009).

La popularidad del tagging social se debe a que ofrece la posibilidad de participar mediante la adición de palabras clave a usuarios no expertos (Mathes, 2004; Voß, 2007), permitiendo la creación de grandes recursos cuando se cuenta con un número suficiente de colaboradores voluntarios y una meta común (Voß, 2007).

#### 1.3.1. Tipos de etiquetado

El estudio y análisis de los patrones de uso en Delicious y Flickr llevó a Cañada (2006) a diferenciar cuatro estilos de etiquetado, basados en las diferentes motivaciones de los usuarios al asignar las etiquetas, con diferentes consecuencias en el sistema respecto a la utilidad social.

- Etiquetado egoísta: en este tipo de etiquetado la motivación es completamente personal, por lo que el usuario crea tags significativas para él, pero carentes de significado fuera de su contexto personal. De esto se deriva un beneficio social variable, ya que depende de la madurez del usuario que asigna las tags. Cuando usa términos más personales, el etiquetado egoísta llena de anomalías y ruido el sistema. Cuando el usuario empieza a sistematizar el etiquetado (usando más tags temáticas o adjetivas y menos tags de acción) aumenta el beneficio social.

Ejemplos de etiquetado egoísta: *leer, para sara, para clase*

- Etiquetado amiguista: es este caso la motivación es alta, ya que sirve para compartir y para reforzar el sentimiento de comunidad entre un grupo. De este modo, las tags que se asignan son conocidas sólo por ese grupo en particular. El beneficio social es alto,

---

<sup>8</sup> Este concepto puede nombrarse de varias formas: *tagging social, tagging colaborativo, etiquetado social y etiquetado colaborativo*. En este trabajo se usa el término *tagging social* en inglés por ser el más extendido.

especialmente para los grupos pequeños, aunque crece si un tema minoritario pasa a ser de dominio público, ya que esas tags pasan a ser conocidas.

Ejemplos de etiquetado amiguista: *reunión del sábado 16, magisterio promoción 2012*

- Etiquetado altruista: este es el tipo de etiquetado donde la motivación es más baja, debido al trabajo que conlleva y a que no hay ningún beneficio directo asociado para el generoso usuario que lo realiza. Ya que el objetivo final es describir el objeto de la manera más útil para que sea fácilmente recuperado por otros usuarios, las tags asignadas son descriptivas, conocidas y generalmente aceptadas. Su beneficio social es, por tanto, muy alto, al facilitar la recuperación de la información.

Ejemplos de etiquetado altruista: *moda, años80, rolling stones*.

- Etiquetado populista: en este caso la motivación del taggeador es alta, ya que todos sus esfuerzos están encaminados a que el objeto sea recuperado, para lo que recurre a tags llamativas, populares y subjetivas. En la mayoría de los casos el objeto taggeado es simplemente spam, por lo que el beneficio social es escaso o nulo.

Ejemplos de etiquetado populista: *top10, sexo, muy-bueno*

Estas categorías de etiquetado no son excluyentes, es decir, no es que un usuario siempre etiquete de una forma concreta. El etiquetado depende del momento, del esfuerzo que realice el usuario, y del objeto a etiquetar.

#### 1.4. La Folksonomía

El origen del término se atribuye a Thomas Vander Wal y proviene del inglés Folksonomy derivado de *Folk* (popular), que proviene del alemán *Volk*, más *Taxonomía*, del griego *Taxis* (clasificación) más *nomos* (ordenar, gestionar). Empieza a extenderse el término cuando el acto de crear una tag en un sitio web se convierte en una acción sistemática, realizada por un colectivo suficientemente amplio. Esto da lugar al tagging, obteniendo como resultado un vocabulario o conjunto de palabras ensambladas, es decir, que mantienen un ensamblaje orgánico e informal de terminología relacionada, (Vander Wal, 2005b) al que se llama folksonomía.

Atendiendo al resultado una folksonomía es el vocabulario colectivo resultante del proceso del tagging (Trant, 2009). Sin embargo, una definición más compleja (Díaz et al, 2009, 2) lo considera “una forma de aprovechar el conocimiento de las personas de manera organizada y democrática al ordenar y clasificar la información que circula por Internet sobre la base de un ambiente de construcción colaborativo, a través de acuerdos que conducen al logro de un mismo objetivo”.

Este nuevo fenómeno aparece y se difunde gracias a estar ubicado en un contexto sociotécnico (Trant, 2009) propiciador del tagging colaborativo. No se trata solamente de la



disponibilidad de tecnologías (hardware y software) que hacen posible la inmediatez en la transmisión de la información y la simultaneidad. Es, además, el desarrollo de una nueva forma de relacionarse y de atribución de valor a la comunicación, lo que se denomina la “sociedad conectada”. Esta nueva realidad ha supuesto la modificación del rol del usuario, desde un consumidor pasivo de información propio del modelo anterior (Web 1.0) a que desempeñe un papel marcadamente activo, no solo en el consumo sino también en la producción de la información, lo que Rodríguez (2009, 2) llama usuario “prosumidor”. Ya no depende de que profesionales de la información clasifiquen y preparen la información para que el usuario pueda recuperarla. Es el propio usuario quien gestiona la información, con sus propios criterios, usando las herramientas disponibles y, en general, con espíritu de colaboración con otros usuarios.

#### **1.4.1. Características de las folksonomías**

Su característica más destacada se refiere a su generación, al ser creada gracias a la participación de los usuarios (Díaz et al, 2009), quienes, con espíritu colaborativo y sin interés propio, generan y clasifican todo tipo de información - desde enlaces, imágenes, vídeos, música, hasta noticias o artículos especializados - para facilitar el acceso a otros usuarios. Este proceso acumulativo redundante en ventajas para todos los usuarios que crecen cuando lo hace la agregación (Weiss, 2005).

La organización de las folksonomías coincide con la de un sistema colaborativo (Cobo y Pardo, 2006) cuyo funcionamiento se basa en la clasificación mediante tags (Díaz et al, 2009). Las tags son elegidas por los usuarios libremente para describir los recursos que se comparten y, siguiendo el espíritu abierto de las folksonomías, todos los usuarios pueden crearlas sin problemas ya que se requieren muy pocas reglas (Weiss, 2005).

La falta de normas al elegir las tags da como resultado la formación de un vocabulario no controlado, se trata, por tanto, de un sistema de categorización no jerárquico, donde la información se presenta en forma de conjuntos, no de jerarquías, que no se rompe cuando hay datos incompletos (Weiss, 2005). Estando de acuerdo sobre la falta de jerarquía entre las tags, Mathes (2004) especifica que, aunque no haya relaciones explícitas y normalizadas entre ellas, sí se crean de forma natural agrupaciones de tags relacionadas, como en el caso de las que comparte una misma URL.

Por su parte, Moreiro (2006) hace hincapié en la retroalimentación, un elemento introducido por las folksonomías que no está presente en otros vocabularios. En el momento en que un usuario elige una tag y la asigna a un recurso, pasa a ser visible en el sistema, integrándose en una agrupación de tags. Posteriormente el creador puede suprimir y editar la tag y, si lo desea, restringir su acceso a otros usuarios o grupos.

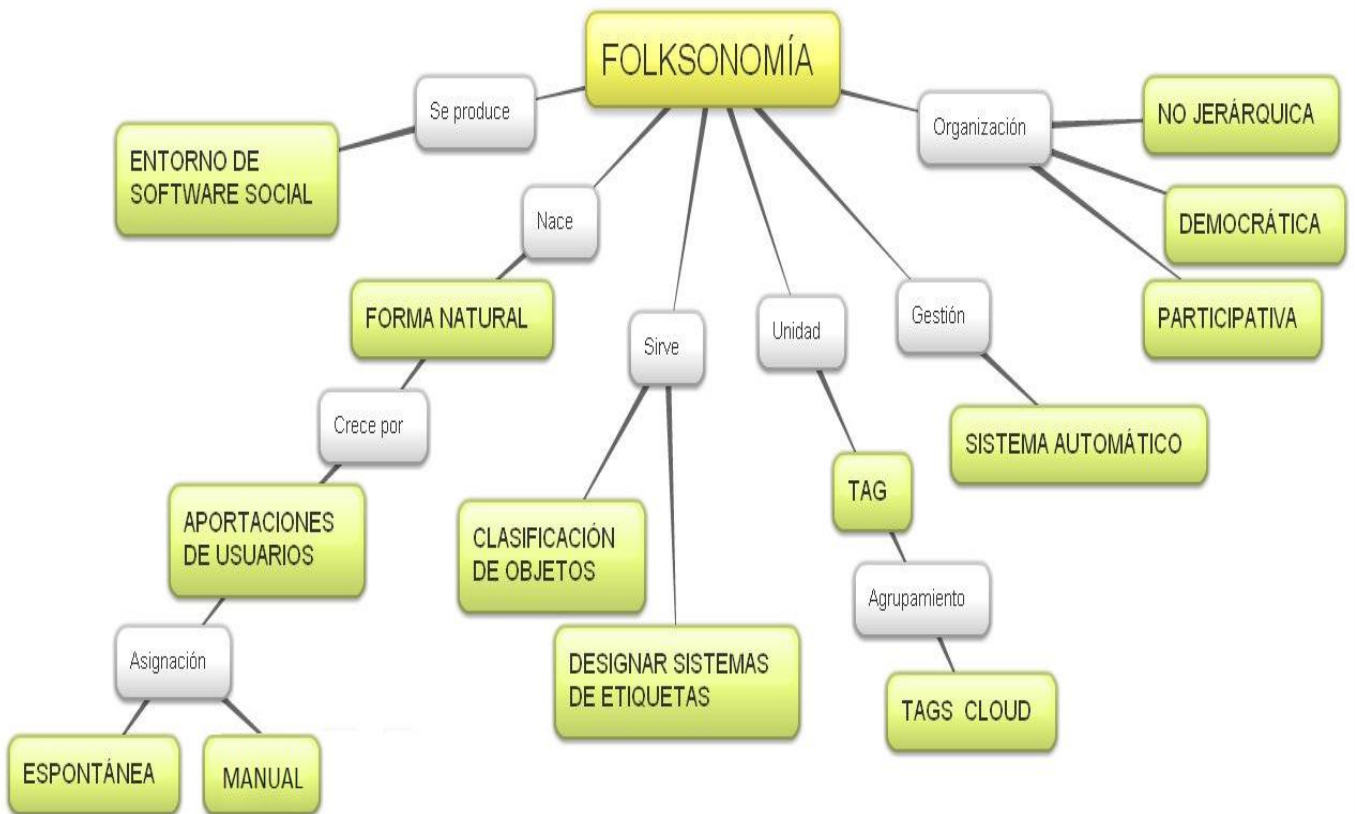


Figura 2. Mapa conceptual de una Folksonomía<sup>9</sup>

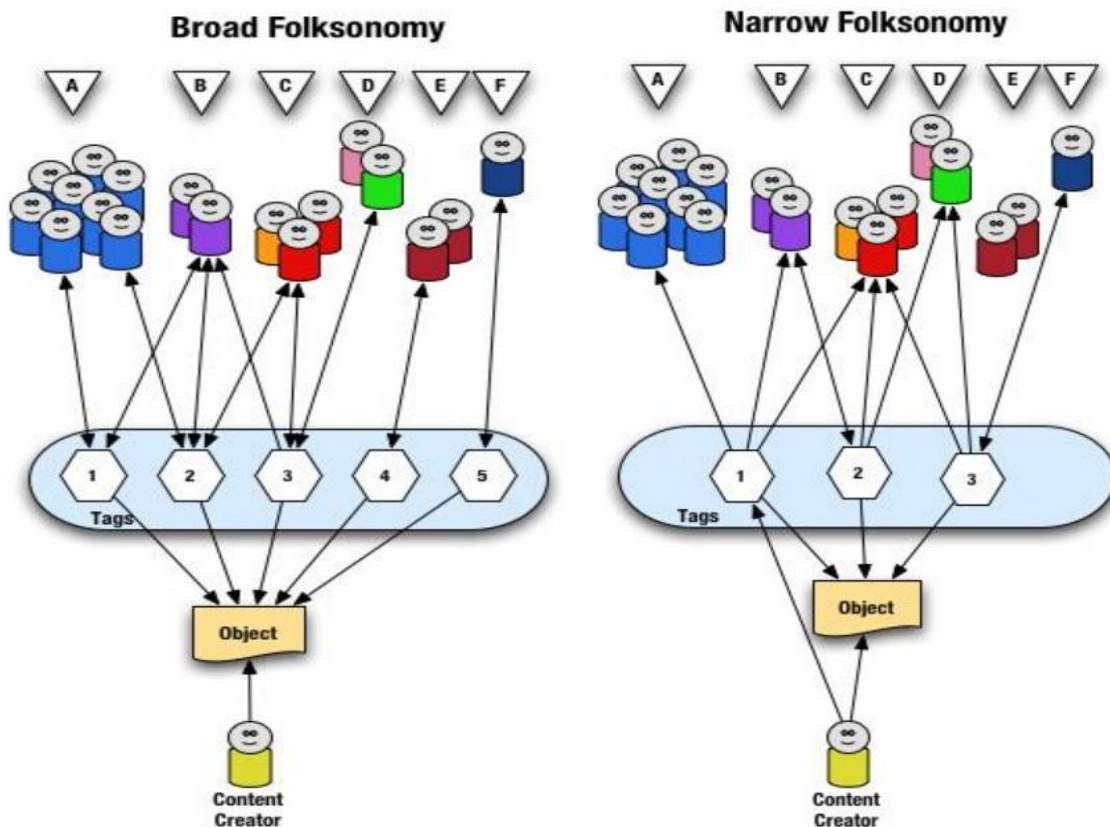
#### 1.4.2. Clasificación de las folksonomías

Se observa una gran aceptación reflejada en las abundantes citas incluidas en la bibliografía que reconocen la aportación de Thomas Vander Wal (2005a) al distinguir dos tipos de folksonomía: amplia y estrecha.

- En la folksonomía amplia muchos usuarios asignan tags a un mismo objeto teniendo cada usuario libertad para asignar tags usando su propio vocabulario. Un ejemplo es Delicious.
- En la folksonomía estrecha una o varias personas<sup>10</sup> designan tags para que la gente recupere esa información posteriormente. Generalmente el creador de contenido asigna una o más tags para animar al resto de usuarios. Un ejemplo es Flickr.

<sup>9</sup> Fuente: elaboración propia

<sup>10</sup> En el documento original, Thomas Wander Val usa la expresión “one or a few people”, es decir, una o pocas personas. En la traducción se optó por la palabra varias, que se considera más adecuada.

Figura 3. Folksonomía amplia<sup>11</sup>Figura 4. Folksonomía estrecha<sup>12</sup>

En estas figuras se observa que en ambas folksonomías aparece la figura del creador de contenido, las tags (representadas por números), y los usuarios representados juntos (A, B...F) por coincidir en la elección de una tag, que asignan tags al recurso y lo recuperan posteriormente (representado por flechas).

Ambas tienen en común que el creador del contenido pone a disposición de los usuarios un recurso (u objeto) al que le son asignadas tags, y que existen usuarios organizados en grupos que participan asignando y/o recuperando el recurso.

Difieren en que en las folksonomías amplias:

- La tarea de asignar tags corresponde enteramente a los usuarios, mientras que en la estrecha el creador de contenido también asigna tags.
- El número total de tags asignadas es mayor.
- La participación de los usuarios al asignar tags es mayor.
- Todos los grupos asignan tags, mientras que sólo dos grupos de la folksonomía estrecha lo hacen.

<sup>11</sup> Fuente: Vander Wal, T. Explaining and showing broad and narrow folksonomies. Off the Top [blog]. Disponible en: <http://www.vanderwal.net/random/entrysel.php?blog=1635> [Fecha de consulta: 3 de enero de 2013]

<sup>12</sup> Idem

- La mitad de los grupos recuperan el recurso usando sólo las tags asignadas por ellos mismos, el resto recuperan las propias y las de otros, mientras que en la folksonomía estrecha, sólo los dos grupos que han asignado tags recuperan las propias, el resto recuperan las tags creadas por otros.

### 1.4.3. Ventajas de las folksonomías

La aceptación e implantación de las folksonomías ha sido rápida y generalizada por toda la Web. Cada vez más autores reconocen las ventajas que aportan, considerando las más destacadas las que se comentan a continuación.

En un sentido práctico presentan la ventaja de su bajo coste de producción (Sinha, 2005), ya que al ser un sistema colaborativo son los usuarios quienes aportan y clasifican la información. Tanto el acceso a la página como el etiquetado es libre, favoreciendo la participación y representación de todo tipo de usuarios, cada uno con su vocabulario propio, su punto de vista y su nivel de conocimiento, no requiriendo ser experto en las materias que clasifica (Mathes, 2004).

En los sistemas de tagging social no es común poner normas al usuario acerca de cómo asignar tags, al margen de recomendaciones superficiales en algunas páginas, de modo que el etiquetado es libre, no siendo necesario aprender un vocabulario específico (Mathes, 2004). A su vez, la asignación de tags tiene un bajo coste cognitivo para el usuario (Sinha, 2005), pues basta con asignar términos que considere que describen el recurso compartido. Además, hay que tener en cuenta que es más sencillo para los usuarios asignar y recordar sus propios términos en vista a una posterior recuperación (Thomas et al, 2009). Por estas razones, Quintarelli (2005) considera que las folksonomías son el reflejo del modelo conceptual de la población, además de responder al lenguaje y a las necesidades reales de los usuarios.

Por otro lado, Shirky (2005) afirma que las folksonomías introducen un elemento para mejorar la capacidad de recuperación en un contexto en el que no es viable la aplicación de lenguajes controlados. En Flickr, un portal que usa el tagging social como sistema de clasificación, los usuarios efectúan las búsquedas con sus propios términos; por ejemplo, si se busca el término "pop", refiriéndose al género musical, el sistema devuelve imágenes de cantantes, conciertos y breakdance callejero. Este tipo de búsqueda libre se opone a la búsqueda realizada en entornos en los que se usan vocabularios controlados, por ejemplo los tesauros. Si en el tesoro de la Unesco se realiza una búsqueda con el término "pop" el sistema no devuelve resultado alguno. Esto no siempre significa que el concepto no este incluido en el tesoro, sino que el término utilizado no es el que se usó para indizarlo. Si se busca el término general música, y se revisan los términos que aparecen debajo, dentro del término genérico "*estilo musical*" aparece el término "*música pop*", como Término Relacionado de "*música contemporánea*", "*música concreta*", "*jazz*", "*música popular*", y "*cultura contemporánea*". Cómo las opciones de recuperación del tesoro requieren que el término que busca el usuario concuerde con el término normalizado por los expertos exactamente, no recuperaremos el descriptor "música pop" si se realiza una búsqueda por el

término pop. De manera Por otro lado, este tesoro adolece de una falta de actualización de sus términos, por ejemplo, no considera el rap ni el hip-hop, contrariamente a los sistemas de tagging social, más representativos de la evolución de la sociedad, de sus modas e intereses.

Las ventajas de los sistemas de tagging social tienden a crecer con la agregación; cuantas más usuarios creadores haya, más rica será la folksonomía. Una repercusión directa de ello es el aumento en la formación de los usuarios en temas específicos (Díaz et al, 2009). Un usuario que empiece con un nivel de conocimiento bajo sobre el proceso de categorizar o recuperar información, a medida que observe cómo están “tageados”<sup>13</sup> otros recursos, descubrirá formas nuevas de usar las tags, sinónimos de los términos, y desarrollará la capacidad de elegir términos generales y específicos más adecuados. Finalmente, el desarrollo del aprendizaje colaborativo y la habilidad de construir nuevo conocimiento redundará en un mayor y mejor aprovechamiento del capital intelectual.

#### 1.4.4. Inconvenientes de las folksonomías

A pesar del éxito de las folksonomías en la Web y de las oportunidades que ofrecen a los usuarios como productores de información, no es difícil encontrar detractores o, simplemente, autores que además de enumerar sus puntos fuertes, también ponen de relieve sus inconvenientes.

El principal inconveniente que se le suele achacar a las folksonomías parte, precisamente, del que es su rasgo distintivo: son los usuarios quienes escogen las tags, sin ninguna autoridad externa que los guíe. El vocabulario resultante, por tanto, es representativo del conjunto de usuarios, y, por ser creado sin seguir normas, es un vocabulario no controlado (Trant, 2009).

Los vocabularios no controlados presentan problemas<sup>14</sup> inherentes a su naturaleza, como la sinonimia, la polisemia, la variación del nivel de especificidad y las anomalías o variaciones léxicas (Thomas et al, 2009). Como consecuencia, se incrementa la ambigüedad de la representación de los conceptos (Díaz et al, 2009), pudiendo llevar a una mala recuperación de la información, restándole valor a la folksonomía.

Estos inconvenientes, causados por los problemas lingüísticos, se verían reducidos si el sistema de tagging tuviera la capacidad de reconocer sinónimos, errores ortográficos y tipográficos, realizar sugerencias, o si permitiera establecer relaciones jerárquicas entre las tags, formando agrupaciones por materias, de forma automática mediante el empleo de aplicaciones de software<sup>15</sup>. O bien, si se cumple la teoría de la autonormalización, con el uso serán los propios usuarios quienes aprendan a regular la producción de tags logrando que sean más precisas y, por tanto, más normalizadas (Trant, 2009).

<sup>13</sup> Traducción libre del término inglés *tagged*

<sup>14</sup> Estos se tratan más detalladamente en el punto 2.4.5. *Problemas semánticos*

<sup>15</sup> En el punto 2.4.5. *Problemas semánticos* se explica el trabajo de un grupo de autores que se centran en el desarrollo de estas aplicaciones.

Por el momento, la opinión más aceptada es que la falta de normas establecidas por los administradores resulta en un conjunto de tags caótico y no controlado que no permite la búsqueda efectiva como un vocabulario controlado (Guy y Tonkin, 2006).

#### 1.4.5. Problemas lingüísticos de las folksonomías

La aproximación a un tema de reciente aparición siempre presenta el inconveniente de encontrarnos con una bibliografía muy limitada. Destaca un grupo de autores (Angeletou et al, 2008; Specia y Motta, 2007; Giannakidou et al, 2008) que focalizan sus esfuerzos, más que en el estudio de los problemas lingüísticos en si mismos, en la búsqueda de soluciones a través del desarrollo de programas informáticos que corrijan de manera automática algunos de los problemas enunciados en el punto anterior. Cada autor se enfrenta a este reto desde una línea de acción diferente.

- Angeletou et al. (2008) pretenden enriquecer las tags de la folksonomía con semántica formal asociándola a conceptos definidos en ontologías online. Para ello crearon el algoritmo FLOR (FoLksonomy Ontology enRichment), que combina el conocimiento de WordNet<sup>16</sup> y de ontologías online. Se aplicó FLOR a una muestra de tags de Flickr provenientes de 226 fotografías, y se obtuvo una precisión del 93% en un 49 % de las tags.
- Specia y Motta (2007) buscan hacer explícitas las semánticas en los sistemas de tags, organizándolos en grupos de conceptos y ontologías parciales. Para ello, combinan estrategias de preprocesamiento, técnicas estadísticas y el conocimiento de ontologías disponibles en la Web semántica (a través de recursos disponibles actualmente). Los resultados preliminares en los conjuntos de tags de Flickr y Delicious demostraron que era posible generar agrupaciones de tags relacionadas correspondiendo a conceptos de las ontologías, e identificar relaciones entre subconjuntos de esas tags.
- Giannakidou et al. (2008) pretenden mejorar el proceso de recuperación utilizando un acercamiento que une grupos de tags relacionadas y fuentes de información sociales, en el cual se tiene en cuenta simultáneamente los aspectos lingüísticos y sociales de las tags. Los resultados son experimentales, pero demuestran la eficiencia del acercamiento propuesto al agrupar tags relacionadas y recursos.

Se han encontrado referencias a la problemática lingüística en artículos de carácter general (García-Silva et al, 2004; Scott y Huberman 2005; Clay Shirky, 2004; Boixadós, 2012) sobre cuestiones relacionadas con las folksonomías o el tagging social, que tratan de manera superficial la lingüística de las folksonomías, destacando cuatro problemas en los que coinciden la mayoría de autores: la sinonimia, la polisemia, la homonimia y el nivel de especificidad.

<sup>16</sup> WordNet es una gran base de datos léxica en idioma inglés, creada en 1980 por George A. Miller, Profesor de Psicología en la *Universidad James S. McDonnell*, psicólogo investigador emérito y senior.

- La **sinonimia** (Del lat. *synonymia*, y este del gr. συνωνυμία) según la RAE<sup>17</sup> se refiere a “un vocablo o expresión que tiene una misma o muy parecida significación que otro”. Es considerado el problema más relevante en las folksonomías (Scott y Huberman, 2006; Boixadós, 2012) ya que la inconsistencia en los términos asignados a las tags dispersa el etiquetado y dificulta la recuperación de la información, por ser difícil para el usuario estar seguro de haber encontrado todos los objetos relevantes. Una manera de resolverlo, o al menos, reducirlo, sería a través de la herramienta de sugerencias de tags, lo que disminuiría su dispersión. Esta solución es conveniente desde el punto de vista de la recuperación de la información, pero también resulta en un empobrecimiento de la folksonomía (Berrueta, 2007). De hecho, Shirky (2004, línea 9-10) no se posiciona a favor de la implantación del control de sinónimos en la Web, afirmando “*que incluso términos relacionados muy estrechamente, como película, largometraje y filme, no pueden ser resumidos en una sola palabra sin perder parte de su significado y su contexto social*”.
- La **polisemia** (De *poli*, muchos y el griego σήμα, significado) según la RAE<sup>18</sup> se refiere a “la pluralidad de significados de una palabra o de cualquier signo lingüístico”. Estas palabras suenan y se escriben igual, pero su significado es distinto. Por ejemplo, si buscamos la palabra “*pluma*” en el Diccionario de la RAE nos encontramos con 20 significados distintos y si buscamos *pluma* en Flickr (figura 5), obtenemos resultados de dos de esos significados (*pluma de ave* y *estilográfica*). Esta abundancia de resultados actúa como ruido enmascarándolos y dificultando el proceso de recuperación de la información.

---

<sup>17</sup> Definición disponible en: <http://lema.rae.es/drae/?val=sinonimo> [Fecha de consulta: 3 de agosto de 2012]

<sup>18</sup> Definición disponible en: <http://lema.rae.es/drae/?val=polisemia> [Fecha de consulta: 3 de agosto de 2012]

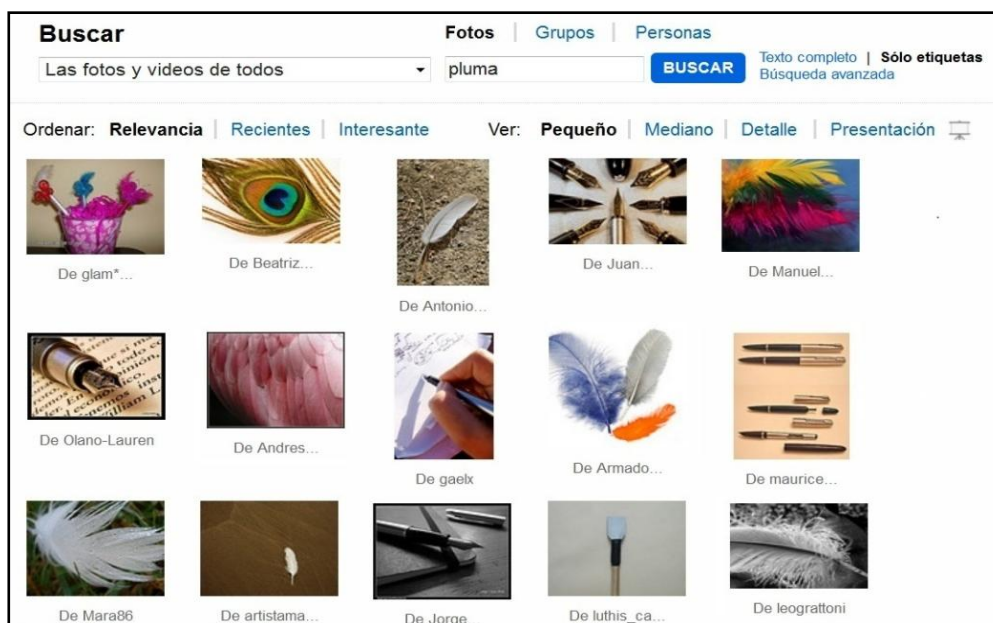


Figura 5. Pantalla de resultados de búsqueda de Flickr

- La **homonimia** (Del latín *homonymiā*, y este del griego ὁμωνυμία, mismo nombre) según la RAE<sup>19</sup> se usa para referirse a “una palabra que siendo igual que otra en la forma, tiene distinto significado<sup>20</sup>”.

Se considera un problema menor que la polisemia, no por su frecuencia de aparición, sino porque existe una solución implementada en algunos sistemas de búsqueda, que ofrecen la opción de indicar qué homónimos se desean descartar en los resultados.

Ejemplo de homonimia: vino (bebido), vino (forma del verbo *venir*.)

Existen palabras homónimas conocidas como **homófonas** que se pronuncian igual aunque se escriben de forma diferente. Por ejemplo: barón (título nobiliario) y varón (ser humano de sexo masculino); bello (que tiene belleza) y vello (pelo corto y suave de algunas partes del cuerpo humano). Es imposible saber si el usuario ha escogido el término adecuado para representar el concepto que quería expresar. Por lo tanto, este tipo de homónimos no se tiene en cuenta en este estudio.

- El **nivel de especificidad** (o variación básica de nivel) se refiere a que los términos relacionados usados para describir un objeto varían su rango de especificidad de muy general a muy específico. Por ejemplo, para un mismo concepto (*gato*), desde un rango general a uno específico, tenemos *animal* > *mamífero* > *felino* > *gato*. Éste último término se considera el nivel básico de especificidad, aquel que relaciona la

<sup>19</sup> Definición disponible en: <http://lema.rae.es/drae/?val=homonimia> [Fecha de consulta: 7 de agosto de 2012]

<sup>20</sup> La diferencia fundamental entre la *polisemia* y la *homonimia* está en el origen de las palabras, es decir, en su etimología. Las palabras *homónimas* tienen etimologías distintas mientras que la palabra *polisémica* tiene un mismo origen, cuyo significado se ha diversificado con el paso del tiempo. Las palabras *homónimas* fueron y siguen siendo palabras distintas que han coincidido en su forma; las llamadas “*palabras polisémicas*” son una sola palabra que ha adquirido distintos significados, entre los cuales hay una relación de sentido.



interacción de los usuarios con el objeto. El problema es que usuarios diferentes pueden considerar términos con distintos niveles más adecuados o apropiados para describir un objeto.

La elección entre los diferentes niveles adoptada por el usuario puede estar relacionada con su nivel de interés, o con el nivel de conocimiento de esa materia (García-Silva et al, 2004). En general, se espera que los usuarios tiendan a usar el nivel básico de especificidad (*gato* en vez de *felino*). Los experimentos realizados por Tanaka y Taylor en 1991 a 16 sujetos, demuestran que, cuando se les pidió que identificaran perros y pájaros en fotografías, los sujetos usaron más los términos “*perro*”, y “*pájaro*”, que “*beagle*” o “*petirrojo*”, y que cuando se les preguntó qué era el objeto de la fotografía, los sujetos respondieron más rápidamente si el objeto era de un nivel básico.

Estos problemas no son los únicos que podemos encontrar. Debido a la falta de normas, las **variaciones morfológicas** están muy presentes en la folksonomía, y contribuyen a la dispersión de las tags. Entre las más usuales: la **conjugación de los tiempos verbales**, la **dicotomía plural/singular**, y la **dicotomía masculino/femenino**. Estas dos últimas causan problemas porque el sistema las almacena como dos tags diferentes sin relación y devuelve resultados truncados. Por ejemplo, si buscamos imágenes de gatos por la tag “*gatos*”, cualquier imagen que haya sido etiquetada como “*gato*”, no será recuperada, por lo que se pierde parte de la información.

Otros problemas aparecen en el momento de asignar tags, ya que en general los usuarios lo realizan a su libre albedrío, sin instrucciones. Uno de los más frecuentes son las expresiones compuestas, una fuente inagotable de problemas. Al usar dos términos (*invitaciones y boda*, por ejemplo), un usuario puede optar por ponerlos juntos o separados. Si los pone juntos, puede diferenciar el inicio de la segunda palabra con mayúsculas o no (*invitacionesBoda*, *invitacionesboda*). La tendencia es a no usar mayúscula al inicio de la palabra, pero la decisión es potestad del usuario. Si decide separarlo, puede introducir cualquier símbolo en medio, como un espacio en blanco, un guión alto o bajo, un punto, una coma...

También se encuentran casos en los que se utilizan símbolos en cualquier parte de la tag que no añaden significado pero responden a algún propósito del usuario. Por ejemplo, en Delicious, hay usuarios que ponen una almohadilla (#) al inicio de la tag, buscando un mejor posicionamiento en la lista de resultados, ya que al ordenarse alfabéticamente los links con la almohadilla aparecerán encabezando la lista.

Hay que tener en cuenta también los **acrónimos** (ya que además pueden tener significados diferentes en distintos idiomas), los **errores ortotipográficos**, y la **dualidad** de algunas palabras que puedan escribirse de dos formas, por ejemplo *oscuro* y *oscura* (este último problema es muy común en el idioma inglés).

Dentro de una folksonomía se puede encontrar tags personales, que sólo resultan de utilidad para sus creadores. Si, por ejemplo, un usuario añade un link que lleva a un artículo y le añade la tag “*para el trabajo de lengua*”, esa tag le facilita su localización, además de darle

un dato importante sobre la utilidad de ese objeto. Sin embargo otro usuario que busque ese artículo, difícilmente utilizará esa tag para buscarlo.

Finalmente, hay que referirse al evidente **multilingüismo** presente en la Web, que causa principalmente dos problemas (Berrueta, 2007). En primer lugar, el uso mayoritario del inglés perjudica a las lenguas minoritarias y a los usuarios con poco conocimiento de dicho idioma. En segundo lugar la existencia de tags que se escriben igual pero tienen significados diferentes según la lengua<sup>21</sup> e incluso pueden llegar a tener significados contradictorios.

Para este trabajo en concreto se han elegido cinco de estos problemas lingüísticos, en base al criterio de resultar más abordables con los recursos disponibles. Estos son: sinónimos, singular/plural, multilingüismo, expresiones compuestas y errores ortotipográficos.

## 1.5. Dos folksonomías actuales: Flickr y Delicious

### 1.5.1. Características de los portales a estudio

#### Flickr:

Es un sitio Web que permite almacenar, ordenar, buscar, vender y compartir fotografías y videos en línea. La condición es que sean fotografías y vídeos creados por los mismos usuarios. Unas normas de comportamiento y condiciones de uso establecidas favorecen la buena gestión de los contenidos.

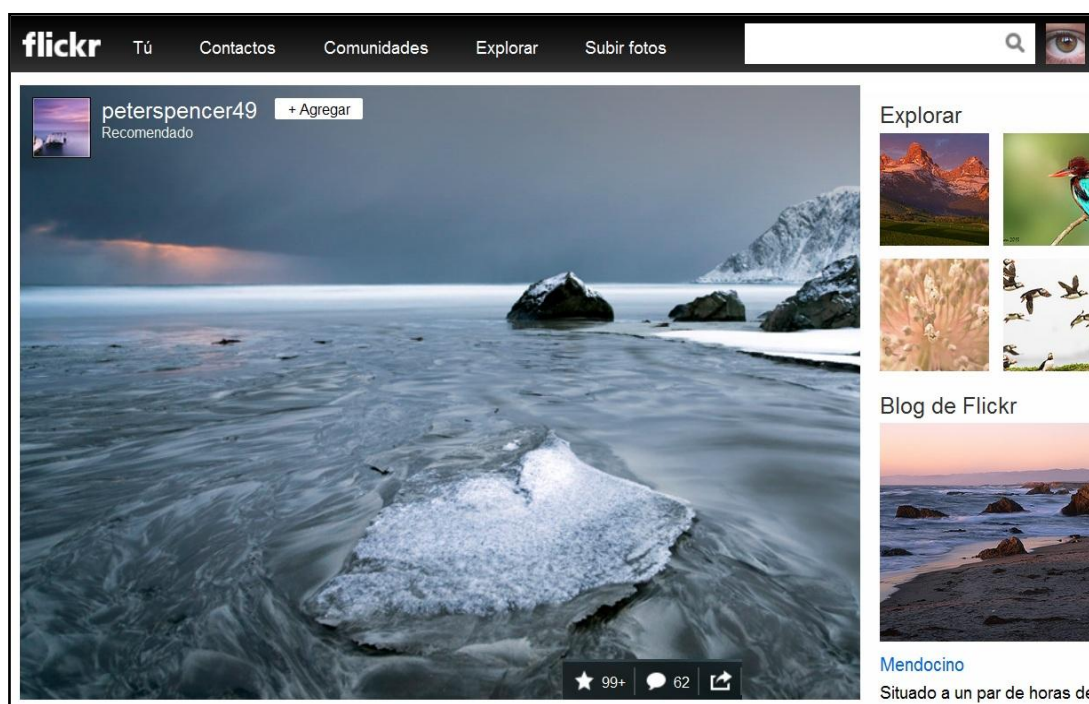


Figura 6. Interfaz básica de Flickr

<sup>21</sup> Por ejemplo, la palabra *once*, en español hace referencia a un número, y en inglés significa “una vez”.

Para publicar imágenes o vídeos es suficiente con registrarse y crear una cuenta gratuita. Esta cuenta tiene ciertas limitaciones, sobre todo respecto a su capacidad, que desaparecen si se adquiere una cuenta PRO. El sistema de Flickr permite hacer búsquedas de imágenes por tags, por fecha y por licencias de *Creative Commons*<sup>22</sup>.

Este portal está muy orientado al ámbito social, permitiendo a los usuarios comentar, votar, guardar recursos, originándose conversaciones y debates entre los usuarios. Al hacer click en un recurso, se abre una nueva página que además de su visualización, permite realizar acciones como escribir comentarios, opiniones, comunicarse con el autor, marcar el recurso como favorito, compartirlo...

La mayoría de los objetos compartidos son imágenes y, en menor medida vídeos, de temas muy diversos. En la comunidad de Flickr hay una gran cantidad de grupos de usuarios unidos por intereses comunes que comparten recursos de alguna temática en particular. En estos grupos los usuarios se ponen de acuerdo para realizar exposiciones temporales, o para, y este punto es muy relevante, etiquetar sus contenidos con tags específicas para facilitar la recuperación posterior por personas del mismo grupo. Se puede decir que en este portal el tagging está orientado a la recuperación posterior por parte de otros usuarios. Muchos se consideran artistas, e intentan que sus contenidos lleguen al mayor número de destinatarios posibles, que pueden votarlos como favoritos y dejar comentarios sobre la imagen.

El portal ofrece pautas y ejemplos pensados para orientar a los nuevos usuarios, o a quienes no tengan claro el proceso de etiquetado, creadas por un administrador en uno de los apartados destinados a la comunicación entre usuarios, el "panel de debate".

Como se ha comentado en el punto 2.4.2. *Clasificación*, Flickr pertenece a lo que Vander Val (2005) llama *folksonomía estrecha*, en la que sólo el creador del contenido o un número reducido de personas aplican las etiquetas al contenido. También hay limitaciones con respecto al número de tags que se pueden asignar, aunque se da la opción de editarlas para crear nuevas y borrar las que el usuario desee.

Durante el proceso de asignación de tags, aparece una explicación básica como orientación: "*Puedes darles a tus fotos y videos una "etiqueta", que es como una palabra clave o identificación de categoría. Las etiquetas te ayudan a encontrar fotos y videos que tienen algo en común. Puedes asignar hasta 75 etiquetas a cada foto o video*". Se pueden asignar las tags que se considere convenientes separándolas por comas. La tendencia es usar el idioma inglés por ser el predominante. Debido a que la mayoría de recursos compartidos son imágenes, se suele describir con tags detalles de la cámara con la que se tomaron las fotografías (la marca y el modelo de la cámara, por ejemplo).

---

<sup>22</sup> Existen 6 licencias *Creative Commons*, cada una con diferentes configuraciones, que permite a los autores decidir la manera en que su obra va a circular en internet, entregando libertad para citar, reproducir, crear obras derivadas y ofrecerla públicamente, bajo ciertas diferentes restricciones. Disponible en: [http://es.wikipedia.org/wiki/Creative\\_Commons](http://es.wikipedia.org/wiki/Creative_Commons) [Fecha de consulta: 21 de noviembre de 2012]

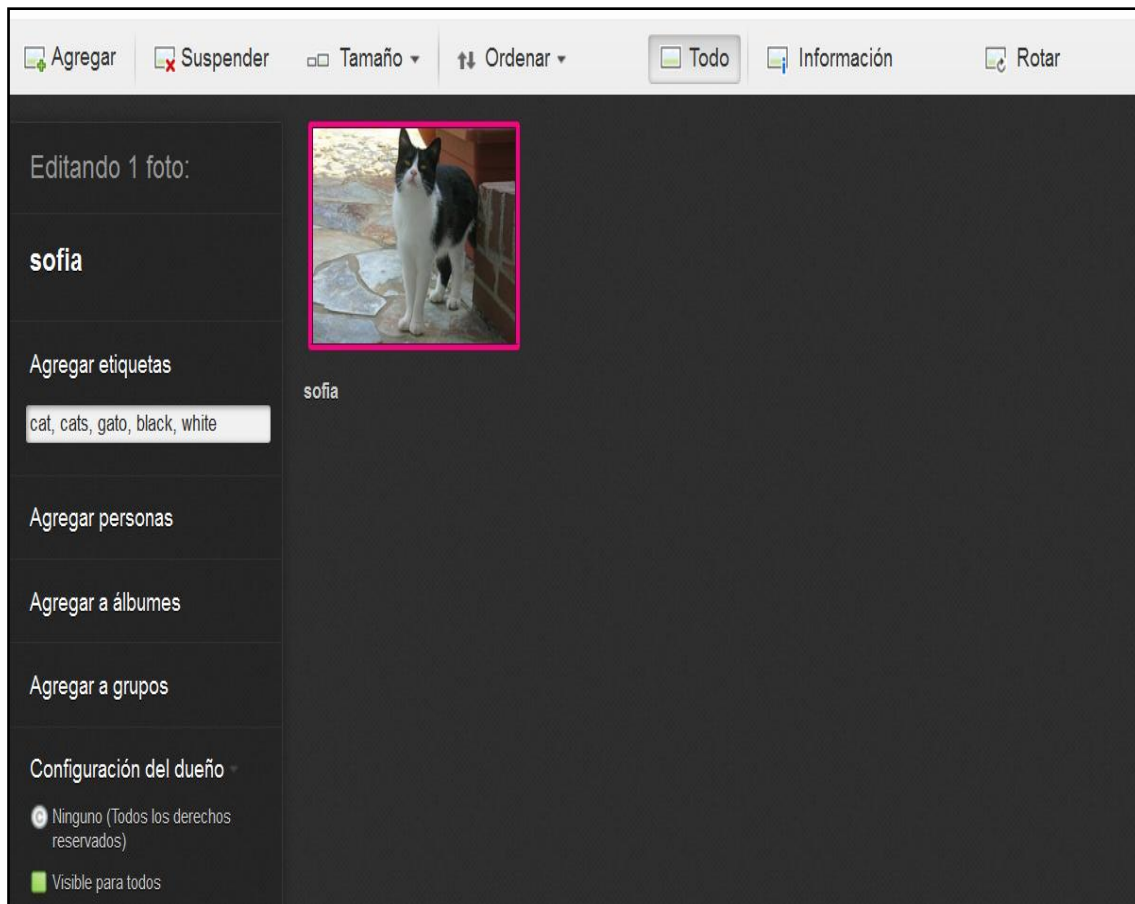


Figura 7. Interfaz de subida de imágenes de Flickr

La página de ayuda a los usuarios (en varios idiomas) ofrece explicaciones sobre la función de las tags, e incluye enlaces a recomendaciones acerca de su elección y redacción, mostrando sugerencias de tags generales sobre temas como formato, tipo de recurso, temática...

### Delicious:

Originalmente conocido como del.icio.us, es un servicio de gestión de marcadores sociales en web. Básicamente permite almacenar, clasificar y compartir enlaces. Inicialmente esta tarea se realizaba guardando las páginas de interés como marcadores (bookmarks) o favoritos en los navegadores. Esto causaba una serie de problemas. El principal es que cuanto más volumen de enlaces había, más difícil resultaba organizarlo y manejarlo, con el inconveniente añadido de que sólo podían consultarse desde el ordenador propio y con frecuencia de perdían debido a problemas técnicos. Para superarlo surgieron los servicios de gestión de marcadores sociales en la Web (social bookmarking).

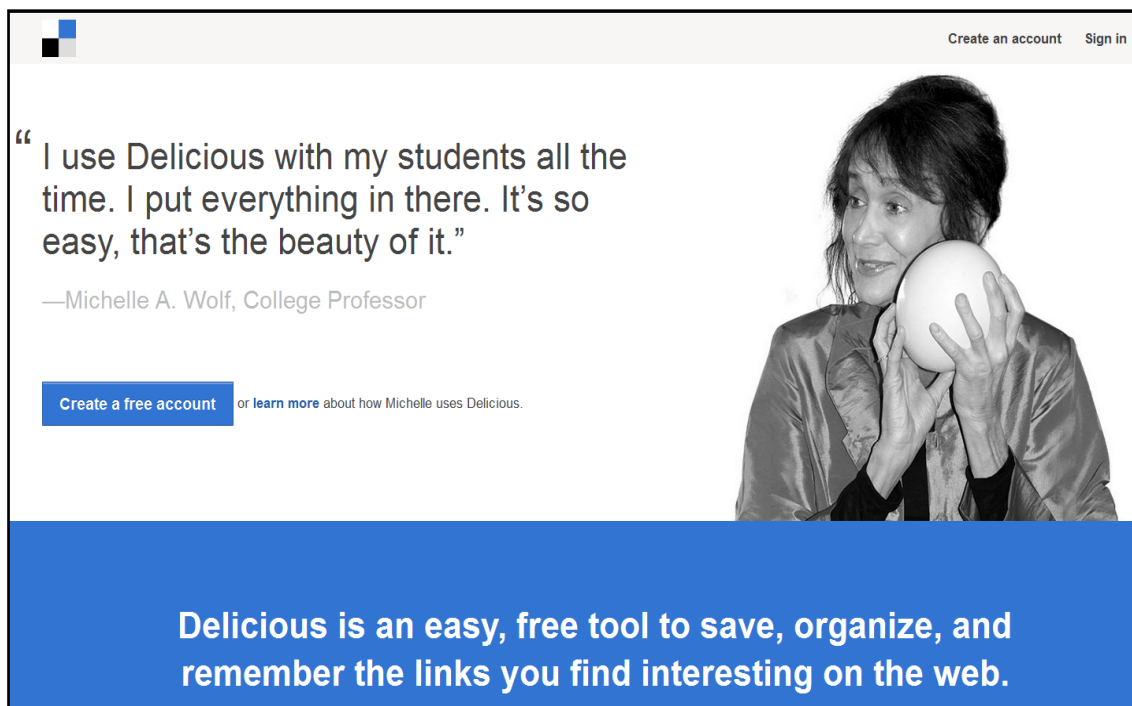


Figura 8. Interfaz básica de Delicious

Permite agregar los marcadores que clásicamente se guardaban en los navegadores y categorizarlos mediante tags para compartirlos con el resto de usuarios, que pueden guardar en su cuenta los enlaces que otros han compartido y asignarles las tags que le parezcan oportunas. También pueden ver lo que otros, con gustos comunes, han compartido o guardado. Inicialmente la búsqueda recupera los resultados por texto, aunque luego permite filtrarlos por tags.

Los enlaces pueden consistir en videos, imágenes, tweets, blog posts, artículos, etc. Siempre que a alguien le parezca relevante, aparecerá. Para facilitar el proceso de subir y guardar links, se permite importar enlaces guardados en la barra de marcadores o favoritos del navegador y también exportar enlaces guardados en su cuenta.

Esta Web pertenece a lo que Vander Val (2005) llama *folksonomía amplia*, donde el creador no influye en las tags que se ponen a su contenido, sino que lo hacen los propios usuarios, favoreciendo que se expresen en sus propios idiomas y con su vocabulario.

Al crear una cuenta, subir un recurso y asignarle tags, no se encuentra ningún tipo de aclaración sobre qué son, para qué sirven, ni tampoco normas establecidas, aunque sí aparecen tags recomendadas (figura 9). Al igual que en Flickr, el usuario asigna las tags que considere convenientes separándolas por comas. La tendencia marcada por las sugerencias es usar el idioma inglés. Su página de ayuda (en inglés), ofrece información básica sobre las tags y su función, pero ninguna recomendación que facilite al usuario la tarea del etiquetado.

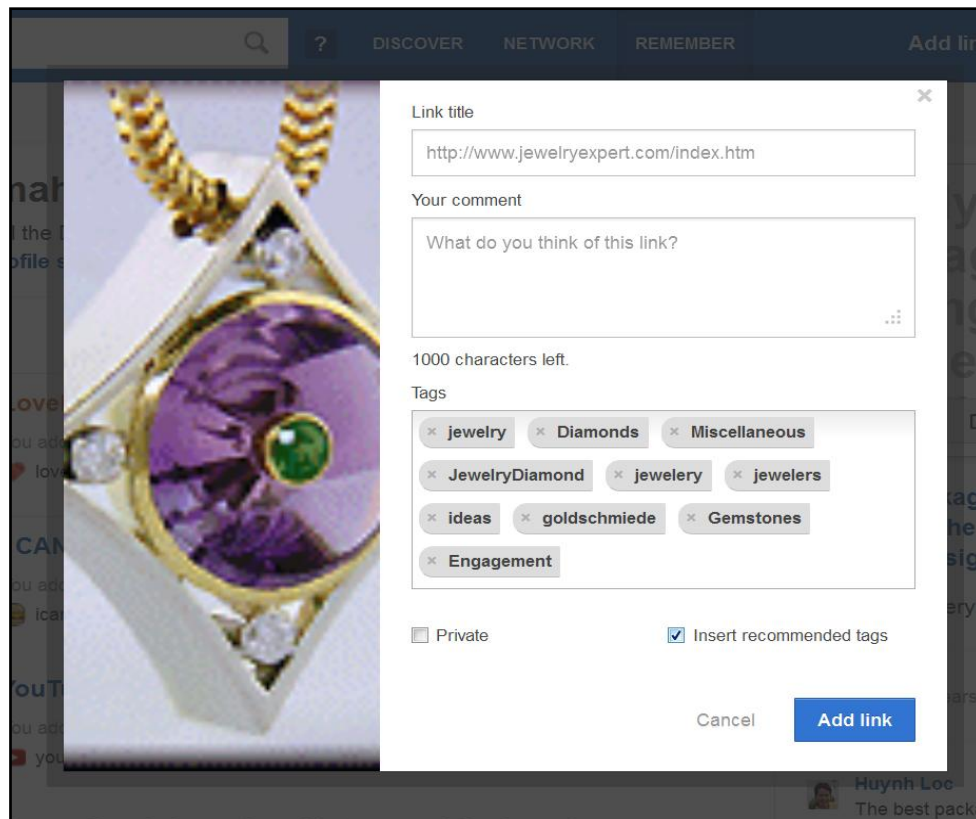


Figura 9. Interfaz de subida de enlaces de Delicious

Las relaciones entre usuarios son menos personales que en Flickr. Básicamente se limitan a seguir a otros y, aunque, hay opción a comentar los recursos que se guardan, la mayoría no comentan nada, o se limitan a exponer con muy pocas palabras si la página les ha resultado o no de utilidad. Parece claro que el etiquetado de Delicious responde más a un interés personal del usuario por organizar sus recursos, lo que lleva frecuentemente a la creación de tags muy específicas sólo útiles para el creador, dándose casos de etiquetados egoístas incorporando mensajes personales sobre los enlaces (*no\_borrar, para\_clase, etc.*).

## 2. Metodología

La primera fase antes de plantear el propósito de este trabajo fue una revisión bibliográfica, ya que sobre el reciente tema del etiquetado social se ha escrito extensamente. La bibliografía utilizada fue una parte recomendación de la tutora y la otra búsqueda personal. Como resultado de esta revisión se concretó el tema del trabajo y su desarrollo.

A continuación se establecieron los objetivos y se llevó a cabo la extracción de datos, siguiendo unas pautas explicadas en profundidad en el punto 2.2. *Técnica de recogida de datos*. Finalmente, estos datos son organizados, procesados y analizados, con la finalidad de obtener la frecuencia de los problemas lingüísticos ya comentados.

### 2.1. Objetivos

Objetivo general:

Cuantificar la aparición de problemas lingüísticos en dos muestras de tags de Flickr y Delicious y analizar su semejanza para verificar si se ven afectados por las mismas variables y con la misma intensidad.

Objetivos específicos:

1. Determinar la frecuencia de aparición de sinónimos y términos equivalentes, tanto intralingüísticos como interlingüísticos.
2. Hallar la frecuencia de aparición cambio del número gramatical para descubrir preferencias por una u otra forma.
3. Determinar la frecuencia de aparición de errores tanto ortográficos como tipográficos.
4. Hallar la frecuencia de aparición de todos los idiomas presentes.
5. Hallar la frecuencia de aparición de expresiones compuestas.
6. Comparar las frecuencias de aparición de cada variable en ambas muestras para determinar las similitudes y diferencias existentes.

### 2.2. Técnica de recogida de datos

Las muestras se tomaron a partir de las *tags más populares* de Flickr y sus relacionadas, gracias a una tag cloud (nube de etiquetas) que agrupa, según aparece en la página web, “*las etiquetas más populares de todos los tiempos*”.

## Las etiquetas más populares de todos los tiempos

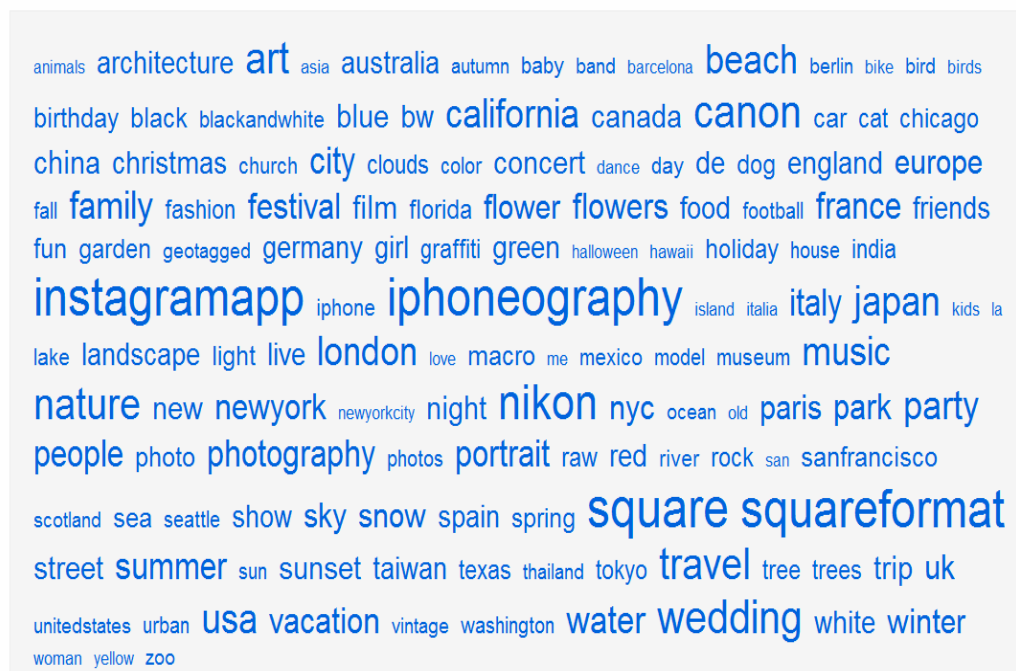


Figura 10. Tag cloud de Flickr

Al hacer clic en cada tag popular, aparecen en una nueva página sus tags relacionadas.

flickr de YAHOO! La visita Crear cuenta Explorar

Explorar / Etiquetas / **art**

Ordenar por:  
 Más recientes • Más interesantes

**Clústeres art**  
 Explora y refina esta lista de art con nuestra maravillosa magia para clústeres.

Etiquetas relacionadas:  
**blue, black, color, light, painting, graffiti, red, green, street, drawing**

Busca cosas similares en

De Serenae De Serenae De takphoto De takphoto

Figura 11. Muestra de tags relacionadas dentro de una tag principal

Se decide trabajar con estas tags populares por ser las más representativas, y de las que resulta más factible extraer una mayor cantidad de datos válidos. Por otro lado, otras categorías como las tags “más populares de las últimas 24 horas”, o “más populares de la



*última semana*” poseen gran inestabilidad, cambian constantemente, y siguen modas pasajeras, por lo que son menos representativas de los términos utilizados por los usuarios.

La organización de Flickr favorece la obtención de las tags relacionadas a partir de las tags principales, a través de la navegación por su propio buscador; aunque no todas las tags principales tienen sus correspondientes relacionadas, ya que se han encontrado 6 tags principales carentes de relacionadas. Las tags vacías encontradas, es decir, las tags principales sin tags relacionadas son: *day, de, la, me, new, san*. Esto no significa que no recuperen resultados (sería difícil que una tag estuviera en la lista de las más utilizadas si no recuperase ningún resultado), ni que una imagen que tenga esa tag no tenga también asignadas otras. Simplemente no hay tags relacionadas que se hayan asignado con la suficiente frecuencia como para resultar relevantes.

La búsqueda en esta web resulta menos laboriosa que en Delicious al poder realizarse sin utilizar la barra de búsqueda, ya que desde la nube de etiquetas se accede directamente a las tags. Sin embargo, en Delicious solo es posible realizarla por medio de su propia barra de búsqueda. Con el fin de verificar el funcionamiento de la barra de búsquedas en Flickr se realizan varios intentos con resultados positivos, quedando comprobado el buen funcionamiento y la correcta recuperación de las tags.

El portal de Delicious no dispone de ninguna tag cloud de tags populares, por lo que se buscan las mismas tags populares de Flickr y las relacionadas de cada una de ellas. El proceso de extracción de datos resulta mucho más complicado que en Flickr, debido al sistema de búsqueda de la página web. Cuando se introduce un término en el buscador, el sistema considera ese término como palabra clave y devuelve enlaces que lo contengan tanto en el texto del documento (título y descripción) como en las etiquetas. En este caso se desea realizar la búsqueda por etiquetas.

Delicious ofrece la posibilidad de buscar por etiquetas, añadiendo el símbolo almohadilla (#) antes del término a buscar. Esta opción no se ofrece en la página de búsqueda, sino que hay que entrar en el apartado de ayuda, ir a herramientas (tools), y buscar entre las preguntas más frecuentes (faq) el apartado *What about searching tags?*, desplegándose un espacio en la pantalla con la explicación para resolver la búsqueda. De este modo se recuperan todas las etiquetas populares y las relacionadas de cada una, que aparecen en la columna de la izquierda. De las 142 tags populares de Flickr se encuentran 136 en Delicious. Entre las 6 que no aparecen, 3 son tags vacías de la lista de Flickr (*de, la, me*).

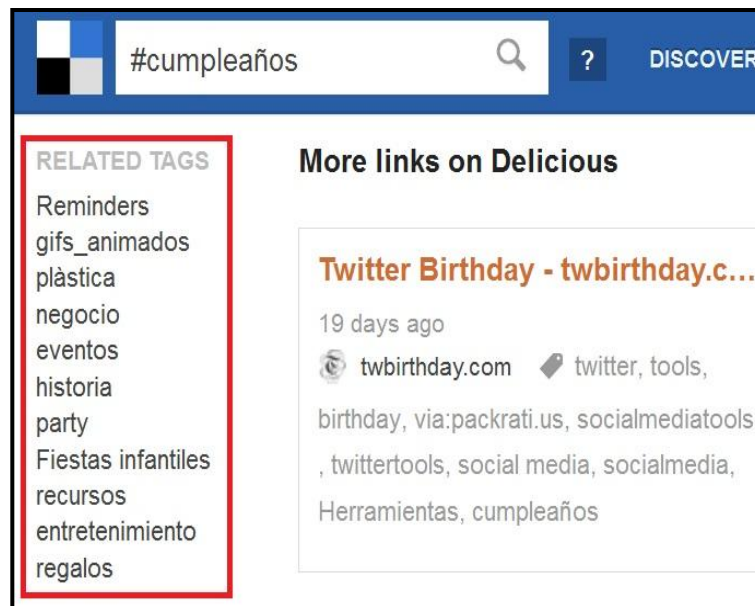


Figura 12. Muestra de tags relacionadas dentro de una tag principal

Conviene aclarar que en el apartado de presentación de resultados se utilizan de manera diferenciada tres términos: término, concepto y etiquetas. El vocablo **término** debe entenderse como la palabra que el usuario escoge para describir el objeto que comparte. **Concepto** es el significado del término, es decir, la idea que representa. Al hablar de **etiquetas** se hace referencia a la ocurrencia, es decir, al número de veces que aparece cada término.

Véase el siguiente ejemplo para ilustrar lo anterior. En la figura 16 se observa que en la columna de la izquierda aparecen los **términos** “cat” y “gato” los cuales, siendo palabras diferentes se refieren a un mismo **concepto** (animal...). En la columna de la derecha aparecen las **etiquetas**, es decir, el número de veces que aparece cada término. Se puede observar que el término “cat” aparece 3 veces y “gato” aparece una vez. De modo que, en este caso, el concepto “gato”, se representa con dos términos, y uno de ellos se ha utilizado más veces que el otro.

	N	O	P
Cat	<b>TÉRMINOS</b>		3
gato			1
Clouds			20
nuvole			2
			<b>ETIQUETAS</b>

Figura 13. Detalle de la muestra de tags

Finalizado el proceso de recogida de datos, se obtiene una muestra de cada portal: un total de 1492 etiquetas en Flickr (las populares más las relacionadas) y, un volumen considerablemente mayor, 12642 en Delicious.

Como puede verse en la figura que sigue, los términos aparecen uno por uno, siendo su valor y el de las etiquetas el mismo. En esta muestra inicial hay términos con varias repeticiones, por lo que se procede a realizar un filtrado agrupándolas y eliminándolas teniendo en cuenta la co-ocurrencia de tag para obtener una muestra limpia que facilite el trabajo posterior.

	A	B	C	D
1	abandoned		abandoned	1
2	air		air	1
3	airme	repeticiones eliminadas y agrupadas	airme	2
4	airme	términos repetidos	america	2
5	america		animal	5
6	america		Animals	2
7	animal		antique	1
8	animal		apple	2

Figura 14. Detalle de la muestra de tags

Tras el proceso de limpieza, el volumen de etiquetas de Flickr desciende un 73,46% y el de Delicious un 61,83%, pasándose a trabajar con una muestra de 396 y 4825 términos respectivamente. Es decir, baja el volumen de términos, pero se mantiene el número de etiquetas.

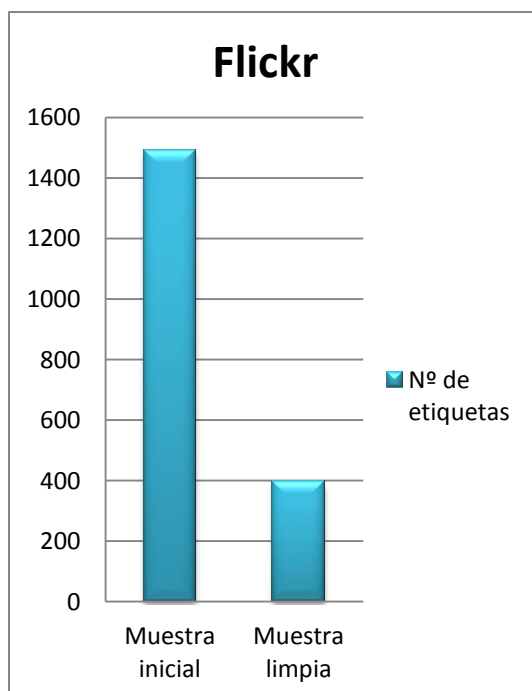


Figura 15. Resultados de la limpieza de Flickr

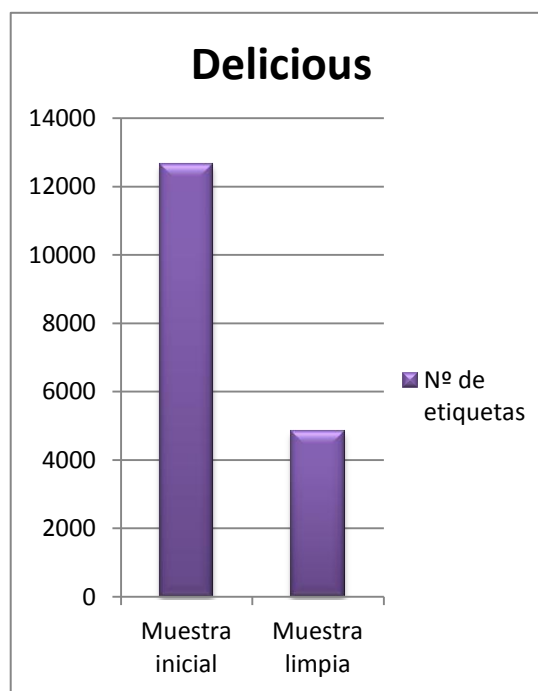


Figura 16. Resultados de la limpieza de Delicious

## 2.3. Métodos de análisis

Una vez se dispone de las muestras con las repeticiones eliminadas y agrupadas, se procede a determinar la frecuencia de los problemas lingüísticos más reconocidos (sinonimia, multilingüismo, dicotomía plural/singular, expresiones compuestas y errores ortotipográficos). Para ello se revisan las muestras, agrupando cada variable en una lista diferente para poder trabajar individualmente con cada una. Posteriormente se calculan los porcentajes de aparición de cada variable en ambas folksonomías. En las variables de *expresiones compuestas*, *multilingüismo* y *plural/singular* se dispone de los resultados de trabajos anteriores, citados en el punto 2.4.5. *Problemas semánticos*, que van a permitir comprobar la concordancia con los obtenidos en este.

**Sinónimos:** se agrupan términos afectados por sinonimia tanto intralingüísticas (del mismo idioma) como interlingüísticas (de diferentes idiomas). Cuando aparece un mismo concepto representado en varios idiomas, por ejemplo “dog” y “perro”, los términos se consideran sinónimos. Se presupone que el valor de aparición de términos afectados por sinonimia debería ser alto, reafirmando la postura de los autores que lo consideran el principal problema de los vocabularios no controlados<sup>23</sup>. En esta variable se agrupan también los términos equivalentes.

Ejemplos:

Sinónimos intralingüísticos: house, home

Sinónimos interlingüísticos: dog, perro.

Términos equivalentes: New York, NY

**Plural/singular:** en esta variable se distinguen dos porcentajes de aparición. Por un lado, se agrupan todos los términos en singular y en plural, con objeto de comprobar cuál de las dos flexiones es más frecuente. Por otro, se agrupan las formas plurales y singulares que se refieren a un mismo concepto (por ejemplo *gato* y *gatos*); en adelante les llamaremos “términos coincidentes”. De esta forma, se averigua si se manifiesta una preferencia de los usuarios por el uso del singular o del plural a la hora de expresar un concepto.

**Errores ortotipográficos:** en este caso se agrupan los errores ortográficos, originados por una confusión primaria en el usuario (no usar tildes, confundir u omitir letras, leísmo...) y los tipográficos, que se producen accidentalmente al escribir en el teclado. Ya que es muy difícil diferenciarlos se consideran un solo tipo de error: ortotipográfico. Es necesario tener en cuenta que para ser considerado un error, debe ser posible identificar el término al que pretende hacer referencia el usuario.

Ejemplo de errores ortotipográficos: fotograafia (fotografía), chickec (chicken), christain (christian), wierd (weird).

<sup>23</sup> Scott y Huberman, 2006; Boixadós, 2012

**Multilingüismo:** se identifican todos los idiomas en que se representan las etiquetas. En las palabras que se escriben igual en varios idiomas, la etiqueta se considera perteneciente al idioma inglés.

**Expresiones compuestas:** entendiéndose como tal cualquier etiqueta compuesta de dos o más términos, independientemente del sistema de separación de palabras.

Ejemplos: `academic_institutions`, `africanmusic`, `animal eyes`, `balance-bike`,  
`blog_food`, `bread.crumbs`, `dessert{fruit}`, `education|research`,  
`filetype:jpg`, `gibbs/dinozzo`, `radio,bbc,english`.

Durante la etapa inicial del trabajo se pretendía analizar también la variable masculino/femenino. Sin embargo, tras comprobar que el idioma mayoritario era inglés (el cual no distingue géneros), se decidió prescindir de dicha variable.

Para los procesos de almacenamiento, procesamiento, análisis y representación de datos se utiliza el programa de hojas de cálculo *Microsoft Excel 2007*. Posteriormente, a través de la estadística descriptiva se llevó a cabo un análisis de los datos extraídos del estudio de estas variables

### 3. Resultados

El proceso de limpieza y clasificación de las etiquetas ha proporcionado datos muy abundantes acerca de la aparición de las variables y su frecuencia. Se presentan a continuación.

#### Flickr

Los resultados para las variables a estudio en el portal de Flickr son:

#### **Sinónimos:**

El número de etiquetas afectadas por esta variable es de 296. Si recordamos que el total de etiquetas de la muestra es de 1492, inferimos un porcentaje de un 19,84% de uso de sinónimos.

Estas 296 etiquetas son representadas por 69 términos, lo que hace una media de 4,29 etiquetas por término. En esencia, se representan 32 conceptos, los cuales son expresados en diferentes idiomas e incluso dentro de un mismo idioma con términos equivalentes.

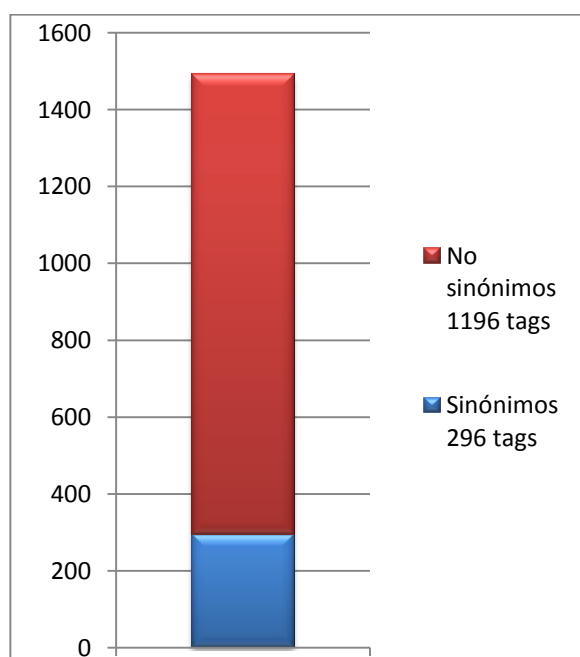


Figura 17. Valor de aparición de sinónimos en Flickr

Dentro del grupo de etiquetas afectadas esta variable aparecen sinónimos en un mismo idioma, otros en diferentes y otros, simultáneamente, en el mismo y en diferentes. Por tanto, en el estudio de esta variable, se distinguen los sinónimos intralingüísticos (un 12,33% de las tags) y los interlingüísticos (un 10,26%), como ilustra la siguiente figura. Puede observarse que

la suma de estos porcentajes no corresponde al total de sinónimos, sino que lo sobrepasa. Esto es debido a que cada término puede estar en un grupo o en ambos al mismo tiempo.

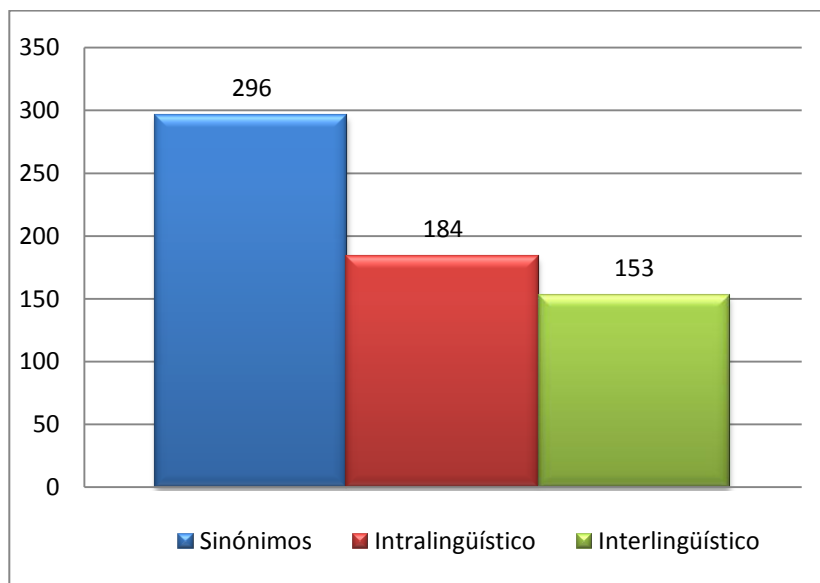


Figura 18. Valores de aparición de sinónimos totales, intra e interlingüísticos

#### Plural/singular:

El número de etiquetas afectadas por esta variable es de 1147. Dado que el total de etiquetas de la muestra es de 1492, se infiere un porcentaje de uso de formas en plural y/o singular de 76,88 %.

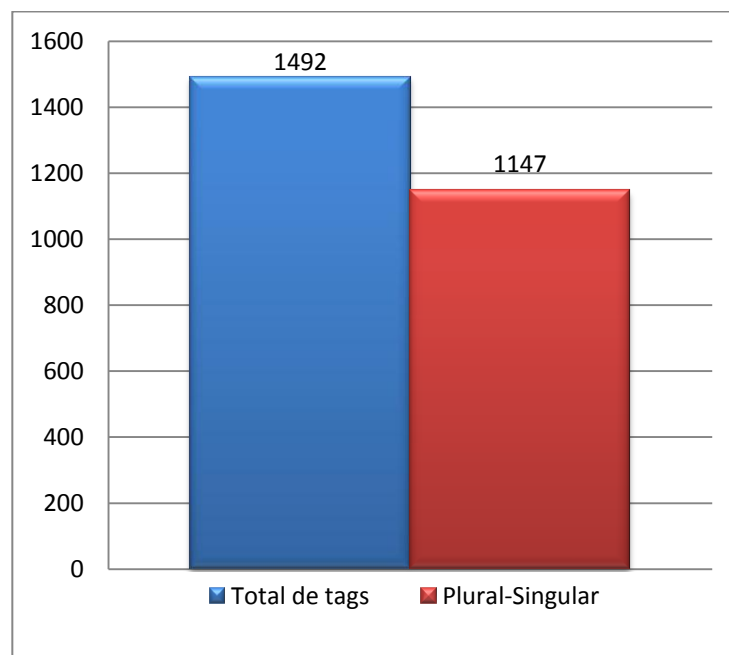


Figura 19. Etiquetas afectadas por la dicotomía plural/singular

En total, 231 términos permiten el cambio del número gramatical, al no ser nombres propios. En los restantes casos no es posible el cambio: nombres de ciudades, países, personas, expresiones populares, nombres de libros, etc.

Dentro del grupo de 1147 etiquetas (231 términos) que lo permiten, un 90,67 % está en singular y el restante 9,33 % en plural.

Como ya se ha comentado en el punto 2.3. *Métodos de análisis*, en esta variable se consideran los “términos coincidentes”, términos que representan un mismo concepto, pero con diferente número gramatical (por ejemplo; gato y gatos).

En la muestra de los 231 términos sinónimos aparecen 185 términos coincidentes, de los cuales el 67,03% está en singular y el 32,99% en plural. Es necesario señalar que estos 185 términos se refieren a sólo 18 conceptos.

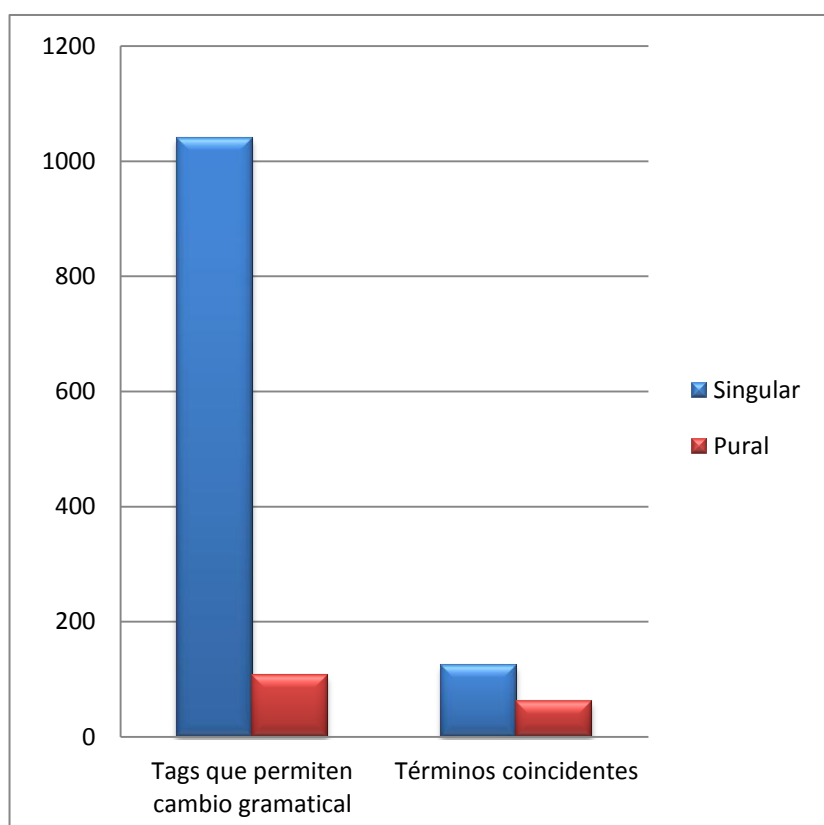


Figura 20. Valores de aparición de la dicotomía plural/singular en Flickr

#### Errores ortotipográficos:

No se encontraron errores ortotipográficos.



**Multilingüismo:**

De las 1492 etiquetas de la muestra, un 96,95% pertenecen al idioma inglés. Aparte del inglés, se identifican cinco idiomas, a continuación enumerados de mayor a menor porcentaje de uso: español, italiano, francés, alemán, catalán.

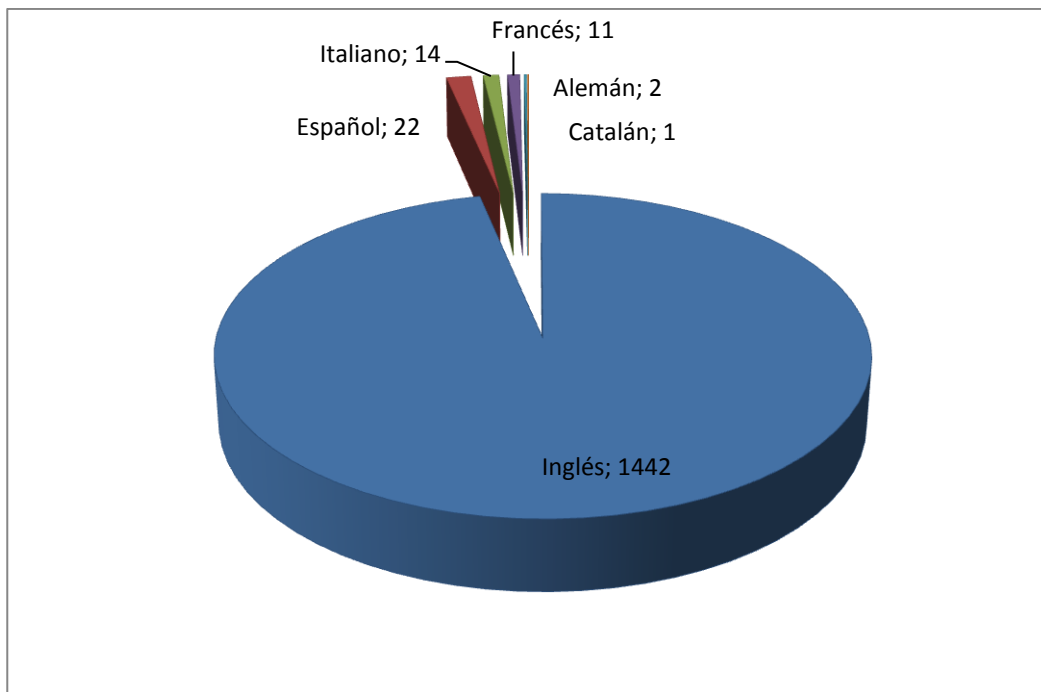


Figura 21. Valores de aparición de los idiomas utilizados en Flickr

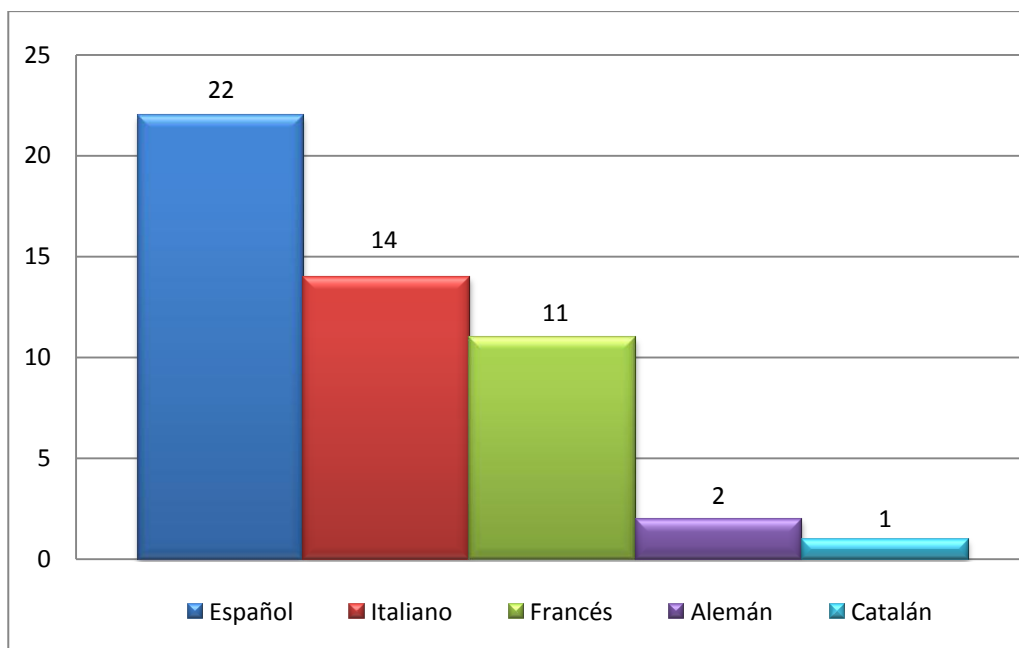


Figura 22. Valores de aparición de los idiomas secundarios en Flickr

**Expresiones compuestas:**

De las 1492 etiquetas de la muestra, un 6% están formadas por más de una palabra.

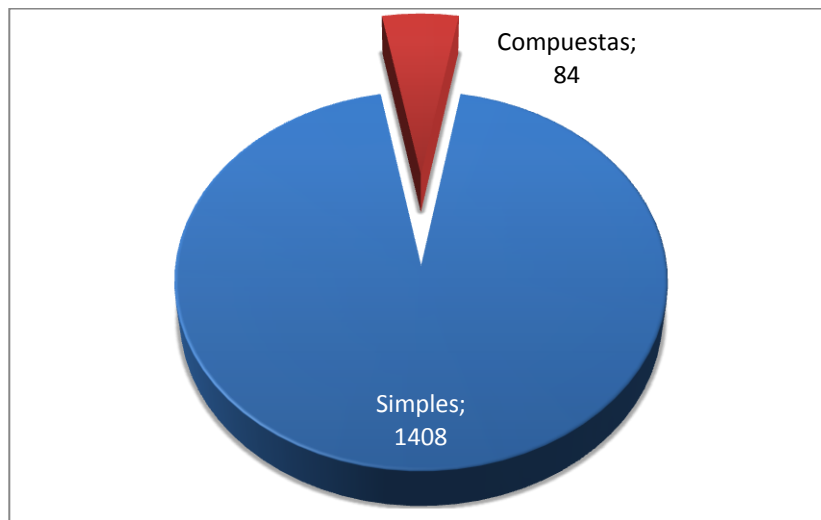


Figura 23. Valor de aparición de expresiones compuestas en la tags de Flickr

**Delicious**

Los resultados para las variables a estudio en el portal de Delicious son:

**Sinónimos:**

El número de etiquetas afectadas por esta variable es de 3862. Si recordamos que el total de etiquetas de la muestra es de 12642, inferimos un porcentaje de un 30,55% de uso de sinónimos.

Estas 3862 etiquetas son representadas por 711 términos, lo que hace una media de 5,43 etiquetas por término. En esencia, se representan 276 conceptos, los cuales son expresados en diferentes idiomas e incluso dentro de un mismo idioma con términos equivalentes.

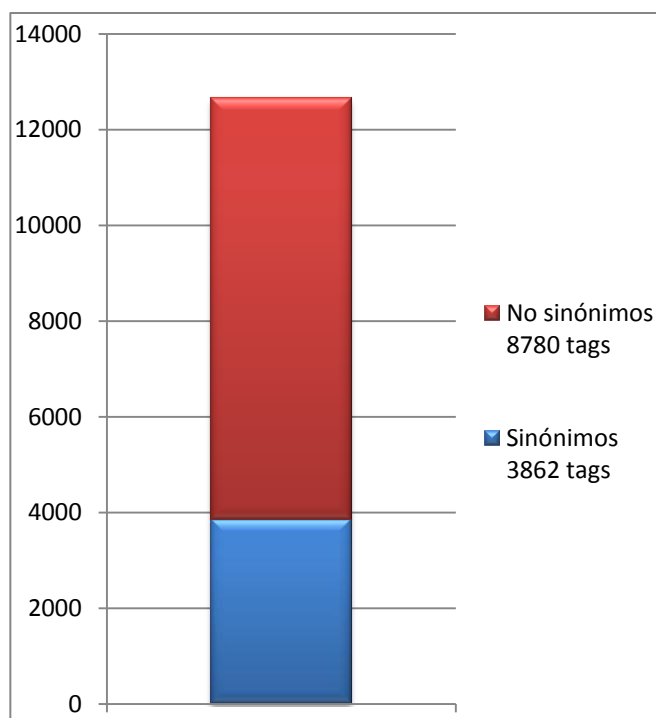


Figura 24. Valor de aparición de sinónimos en Delicious

Dentro del grupo de etiquetas afectadas esta variable aparecen sinónimos en un mismo idioma, otros en diferentes, y, simultáneamente, en el mismo y en diferentes. Por tanto, en el estudio de esta variable, se distinguen los sinónimos intralingüísticos (un 43,66% de las tags) y los interlingüísticos (un 23,04%), como ilustra la siguiente figura. Puede observarse que la suma de estos porcentajes no corresponde al total de sinónimos, sino que lo sobrepasa. Esto es debido a que cada término puede estar en un grupo o en ambos al mismo tiempo.

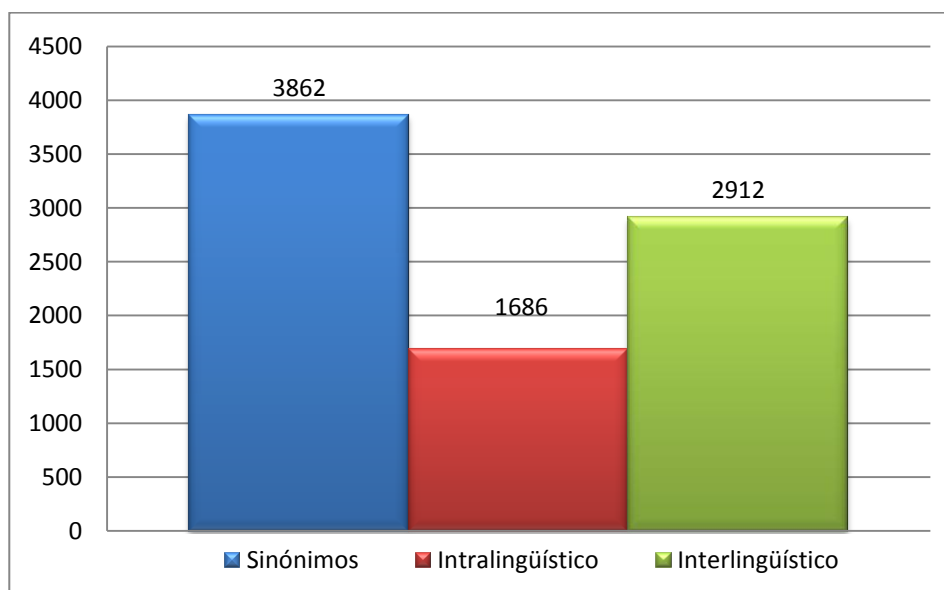


Figura 25. Valores de aparición de sinónimos totales, intra e interlingüísticos

**Plural/singular:**

El número de etiquetas afectadas por esta variable es de 6470. Dado que el total de etiquetas de la muestra es de 12642, se infiere un porcentaje de uso de formas en plural y/o singular del 51,18 %.

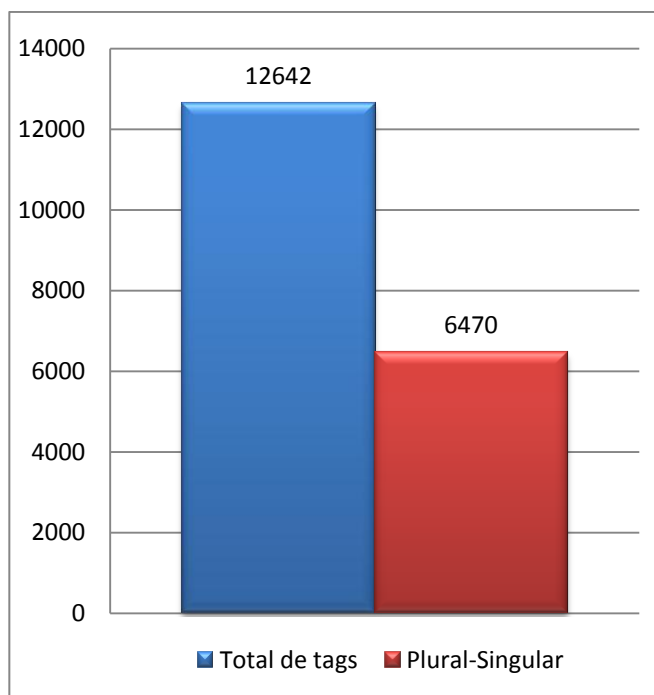


Figura 26. Etiquetas afectadas por de la dicotomía plural/singular

En total, 1448 términos permiten el cambio del número gramatical, al no ser nombres propios. En los restantes casos no es posible el cambio: nombres de ciudades, países, personas, expresiones populares, nombres de libros, etc. Dentro del grupo de 6470 etiquetas (1448 términos) que lo permiten, un 71,19% está en singular y el restante 28,81% en plural.

Como ya se ha comentado en el punto 2.3. *Métodos de análisis*, en esta variable se consideran los “términos coincidentes”, términos que representan un mismo concepto, pero con diferente número gramatical (por ejemplo; gato y gatos).

En la muestra de los 552 términos sinónimos aparecen 185 términos coincidentes, de los cuáles el 61,54% está en singular y el 38,46% en plural. Es necesario señalar que estos 185 términos se refieren a 275 conceptos.

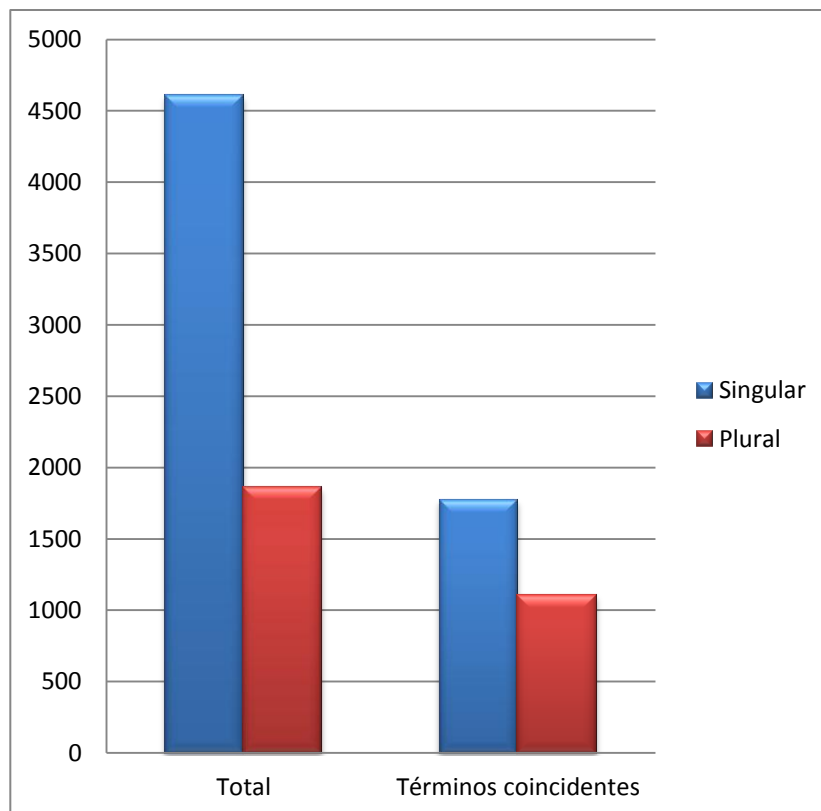


Figura 27. Valores de aparición de la dicotomía plural/singular en Delicious

### Errores ortotipográficos:

De las 12642 etiquetas de la muestra, 91 términos son errores de ortografía, representando un 0,72% del total.

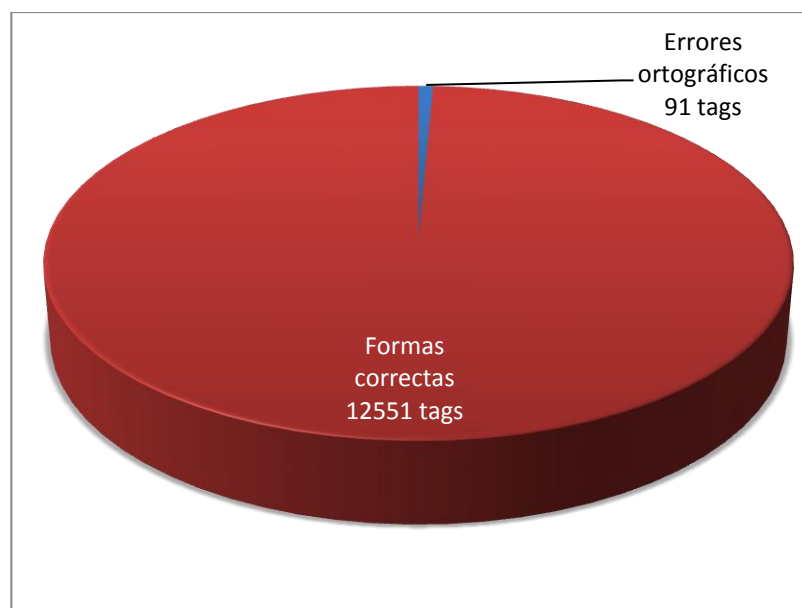


Figura 28. Valor de aparición de errores ortotipográficos en Delicious

**Multilingüismo:**

De las 12642 etiquetas de la muestra, un 94,91% pertenecen al idioma inglés. Aparte del inglés, se identifican 26 idiomas secundarios, de los cuales los 8 de mayor uso suponen el 92,24% de las etiquetas.

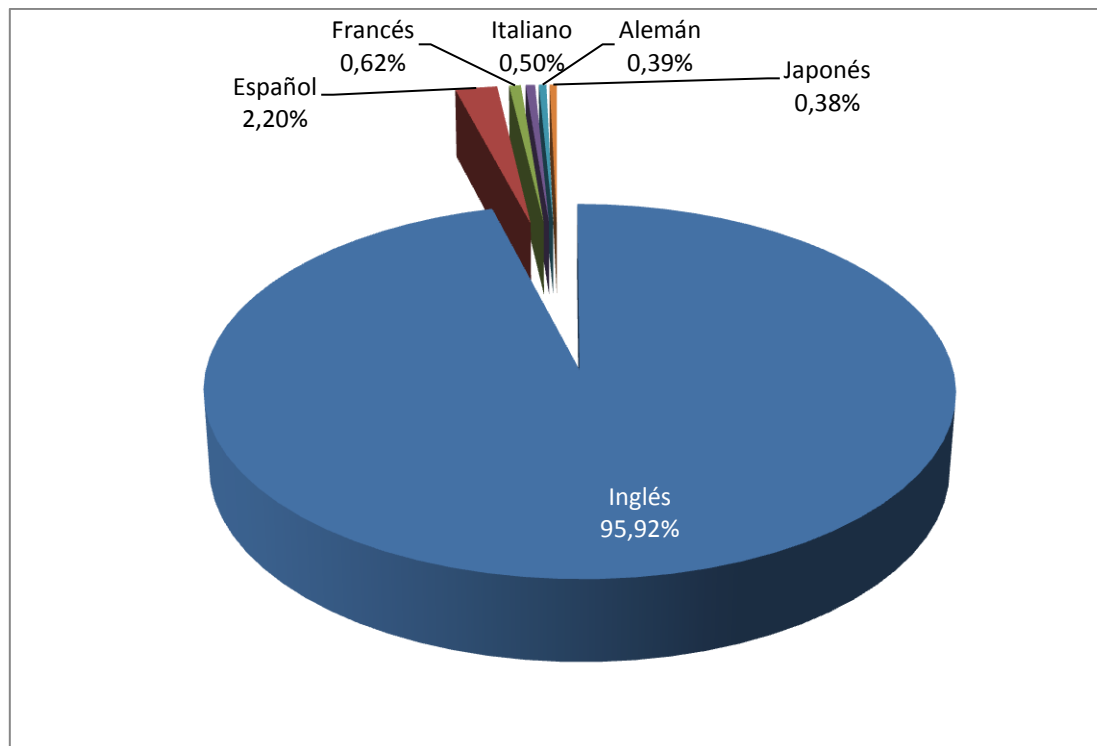


Figura 29. Valores de aparición de los idiomas más utilizados en Delicious

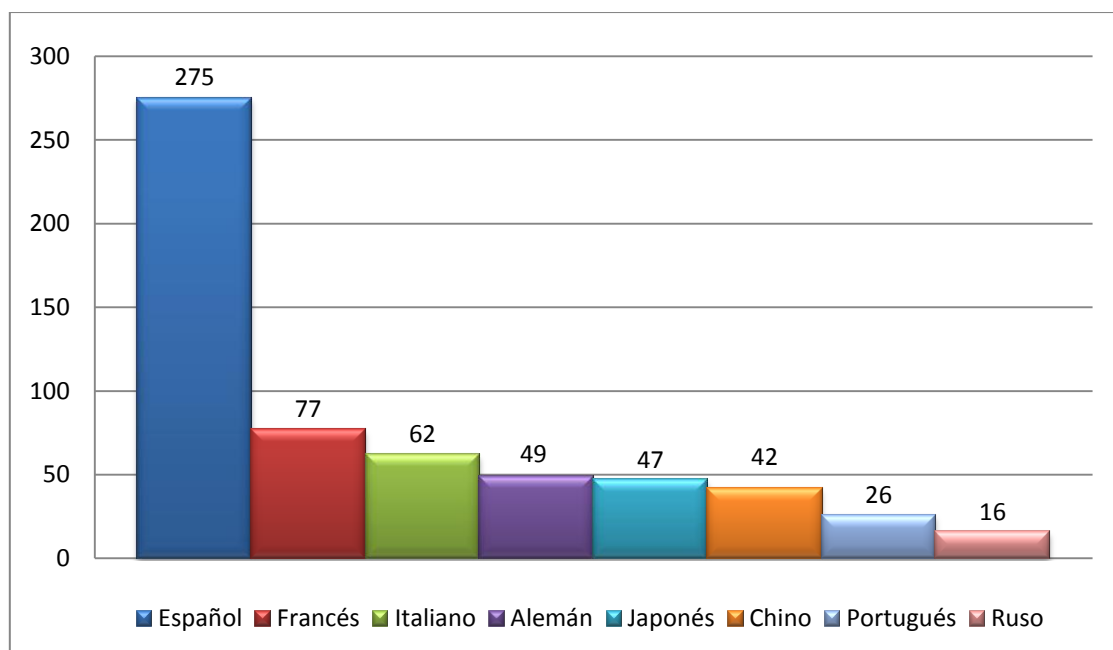


Figura 30. Valores de aparición de los idiomas secundarios más utilizados en Delicious

### Expresiones compuestas:

De las 12462 etiquetas de la muestra, un 16,77% son expresiones compuestas.

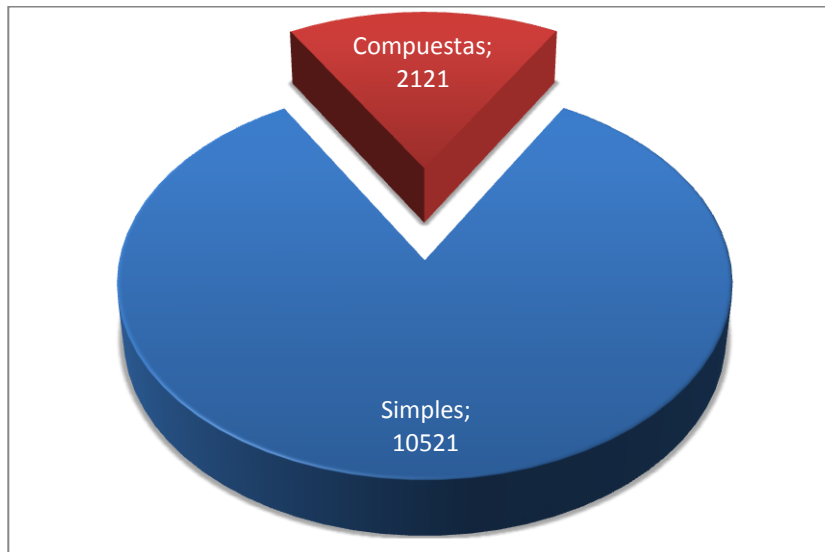


Figura 31. Valor de aparición de expresiones compuestas en la tags de Delicious

A continuación se presentan varias figuras ilustrando las diferencias en los valores de Delicious y Flickr, con la intención de facilitar la comparación entre ambos portales.

### Flickr y Delicious

#### Sinónimos

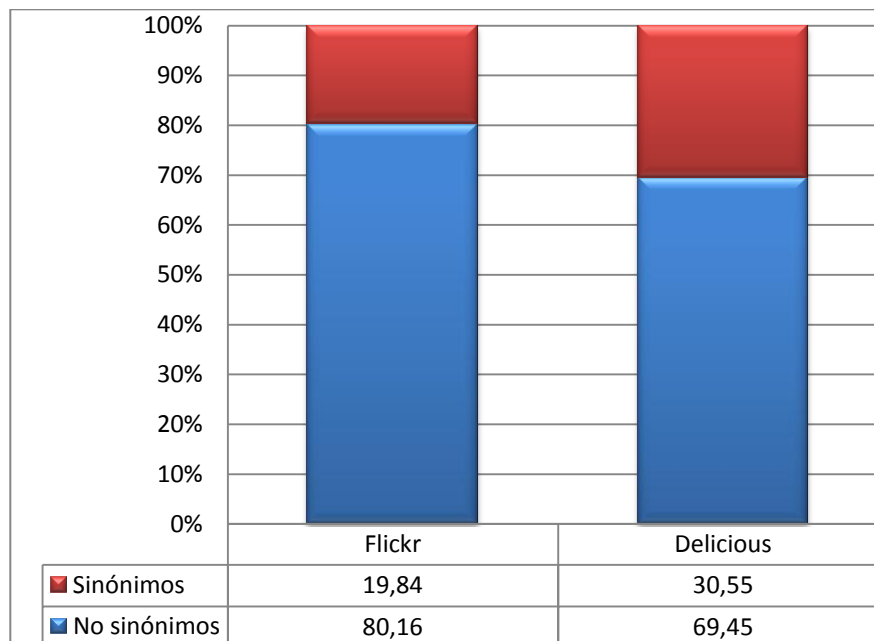


Figura 32. Valores de aparición de sinónimos en Flickr y Delicious

Cómo ya se ha comentado en el punto 2.4.5. *Problemas semánticos* del presente trabajo, la sinonimia se considera el problema más relevante de las folksonomías (Scott y Huberman, 2006; Boixadós, 2012). En ambos casos el valor de aparición de sinónimos es considerado medio (ni muy alto ni muy bajo), siendo de aproximadamente 20% y 30% en Flickr y Delicious respectivamente sobre el total de etiquetas, lo cual es positivo para los usuarios. Cuanto más alto sea la presencia de sinónimos en un vocabulario no controlado, más difícil es la recuperación de información, debido a que el sistema no reconoce relaciones entre las etiquetas.

El hecho de que los valores obtenidos sean medios muestra una tendencia de los usuarios a utilizar las mismas etiquetas para referirse a conceptos similares, lo cual facilita la recuperación de información, además de ser un punto a favor de los autores (Weiss, 2005; Trant, 2009) que confían en la capacidad de los usuarios para regular el vocabulario utilizado sin marcar reglas y normas estrictas.

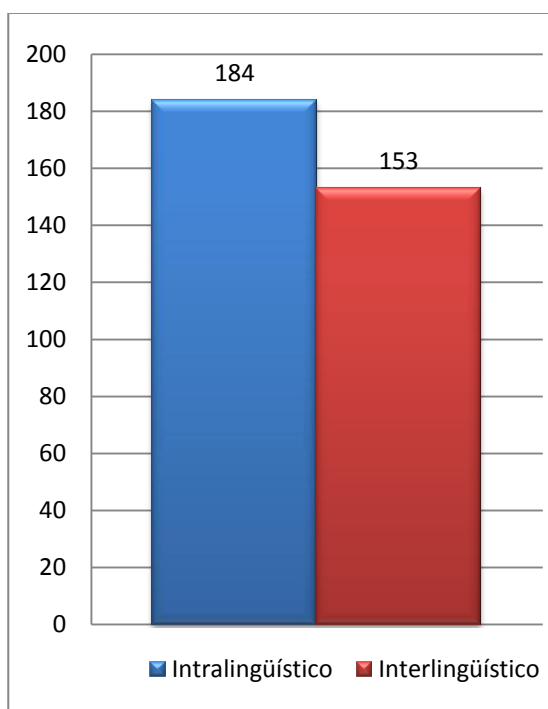


Figura 33. Valores de aparición de sinónimos intra e interlingüísticos

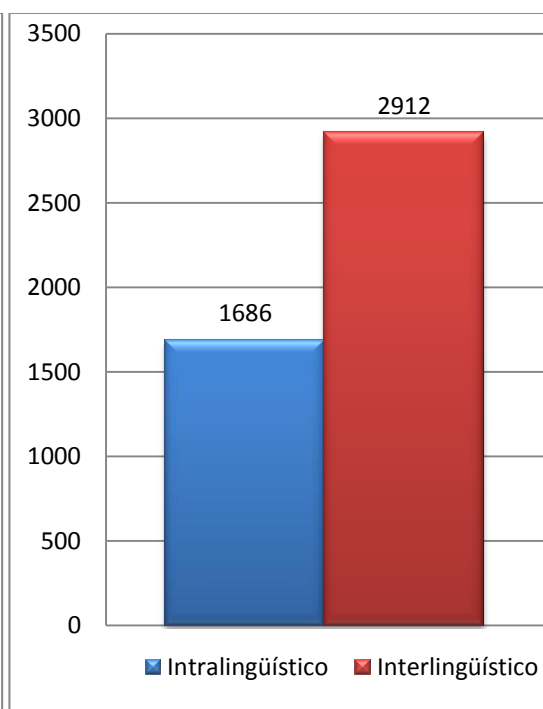


Figura 34. Valores de aparición de sinónimos intra e interlingüísticos

Respecto a los valores de aparición de sinónimos intra e interlingüísticos, es evidente la diferencia entre ambos portales. En el portal de Flickr los valores están bastante igualados, habiendo entre ambos una diferencia de apenas 30 etiquetas, y con una aparición mayor de sinónimos intralingüísticos (en un mismo idioma). En el portal de Delicious la diferencia entre ambos valores es amplia (1226 tags), habiendo una clara superioridad de los sinónimos interlingüísticos.

Esta diferencia resulta lógica atendiendo a la variación en el número de idiomas presentes en cada portal; 6 en Flickr y 27 en Delicious. Cuanto mayor es el número de idiomas haya, aumenta la probabilidad de aparición de conceptos representados en varios de ellos.



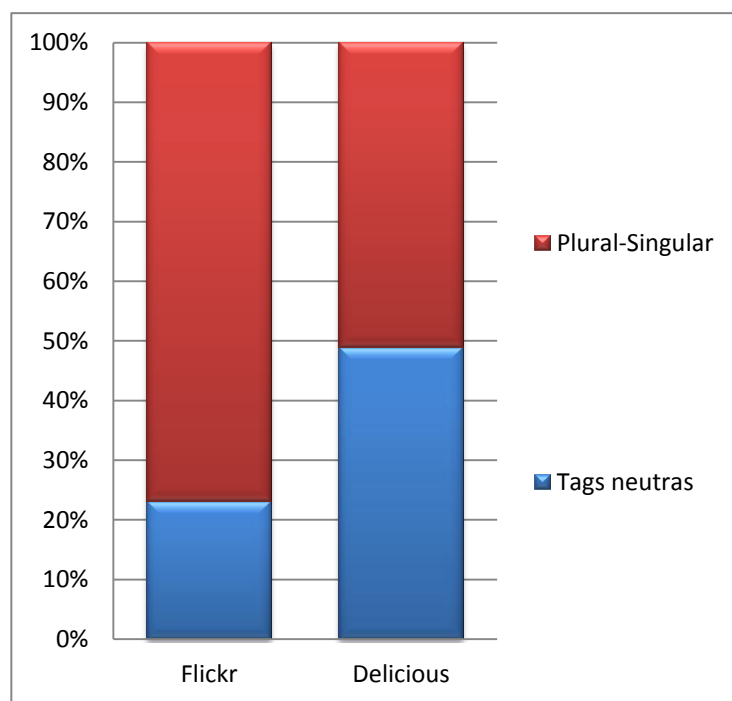
**Plural/singular**

Figura 35. Porcentaje de etiquetas afectadas por de la dicotomía plural/singular en Flickr y Delicious

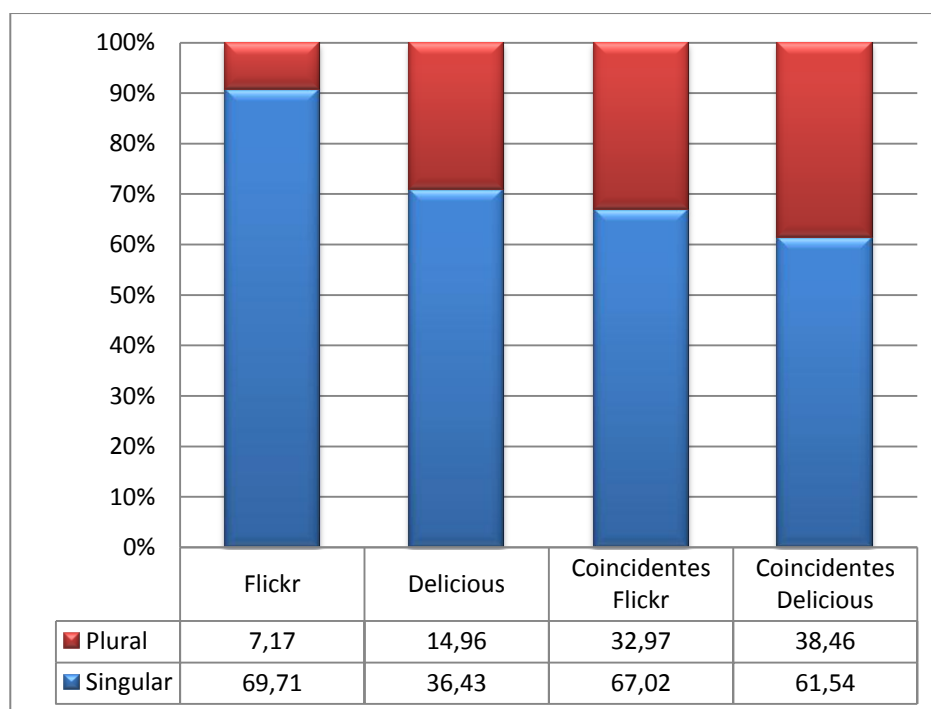


Figura 36. Valores de aparición de la dicotomía plural/singular en Flickr y Delicious

En los vocabularios no controlados es común, debido a la falta de normalización, que un mismo concepto se represente en sus formas singular y plural, lo que añade ambigüedad al vocabulario, ya que el sistema devuelve los resultados indistintamente, perdiéndose parte de la información en el proceso.

En este caso en ambos portales el porcentaje de etiquetas afectadas por esta variable es alto, aunque el de Flickr es superior, casi un 77% de etiquetas afectadas. Tanto sobre el total como en los términos coincidentes se observa una preferencia por las formas en singular, que sigue siendo más marcada en Flickr.

### Errores ortotipográficos

Esta variable no se considera una de las de mayor presencia, aún así el bajo valor de aparición en Delicious sorprende por escaso, apenas 91 tags afectadas. En Flickr el valor es nulo, lo cual es muy inusual, ya que ni siquiera aparecen errores meramente tipográficos.

El ínfimo valor obtenido (0,72%) es positivo para el funcionamiento del sistema y la recuperación de información. Y también, al igual que el bajo valor de aparición de sinónimos (20% y 30%), contribuye a reforzar la creencia en las capacidades de regulación del vocabulario por parte de los usuarios.

### Multilingüismo

Siendo el inglés el idioma más utilizado en la web, no es de extrañar que en ambos portales su uso sea tan alto. La diferencia entre los valores de Flickr y Delicious se entiende si se tiene en cuenta que en el primero se identifican 6 idiomas y en el segundo 27. Entre los lenguajes secundarios comunes más utilizados, en ambos casos el español es el segundo más utilizado.

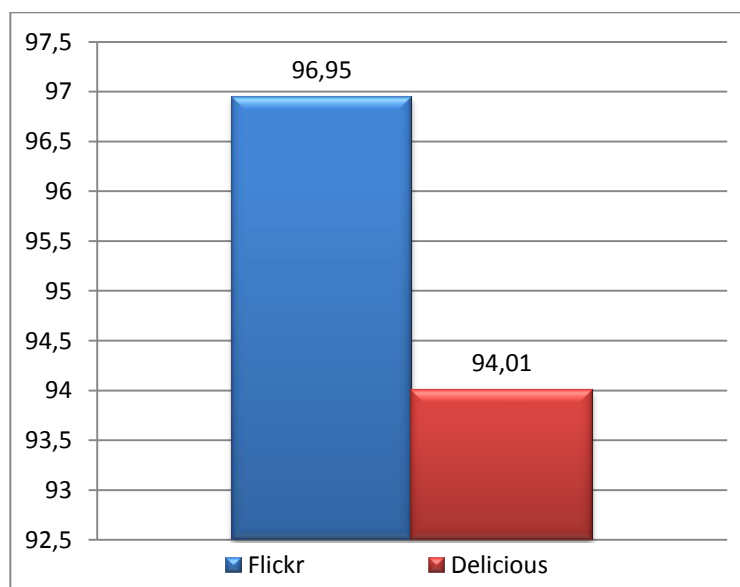


Figura 37. Porcentaje de uso del idioma inglés en Flickr y Delicious

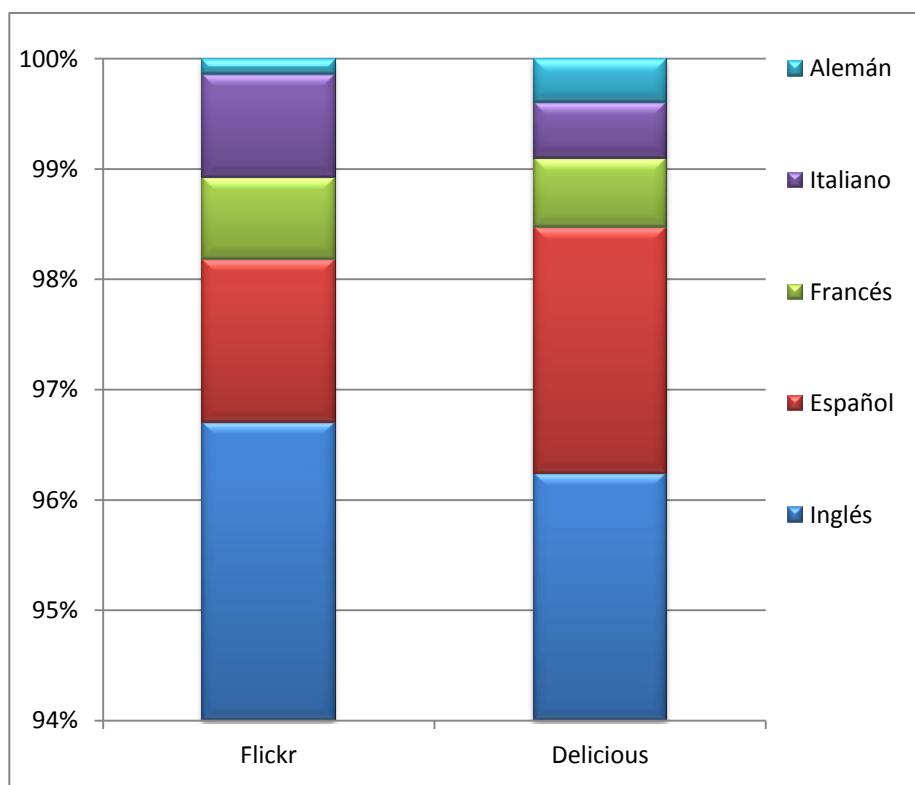


Figura 38. Valor de aparición de los 5 idiomas principales

La presencia de varios idiomas supone un aumento de la ambigüedad del vocabulario, ya que pueden aparecer sinónimos en varios idiomas, o puede no recuperarse información relevante por falta de conocimiento del idioma en que se ha etiquetado. El alto número de idiomas presentes en Delicious conduce a la precitada conclusión de que existe dispersión en el etiquetado. Sin embargo, el análisis de las etiquetas evidencia que el 98,12 % de estas se concentran en los 2 idiomas más utilizados (inglés y español), mientras las 18 últimas lenguas representan el 0,40% de las etiquetas; haciéndose evidente que no existe tal dispersión.

El hecho de que el inglés sea el idioma más usado no es sorprendente, ya que es el más utilizado en la web. Sin embargo, no es posible negar que se produjera un sesgo al seleccionar las tags principales con las que se iba a trabajar. Como ya se ha comentado, estas provienen de la nube de etiquetas “*las tags más populares de todos los tiempos*” de Flickr. De estos 142 términos, un 77,47 % están en inglés, un 2,82 % en español y un 19,72% eran términos equivalentes en varios idiomas (Asia, Chicago, iphone, zoo, nikon...). Al estar la mayoría de las etiquetas principales en inglés, sus relacionadas tienen más probabilidades de estar en ese idioma también.

### Expresiones compuestas

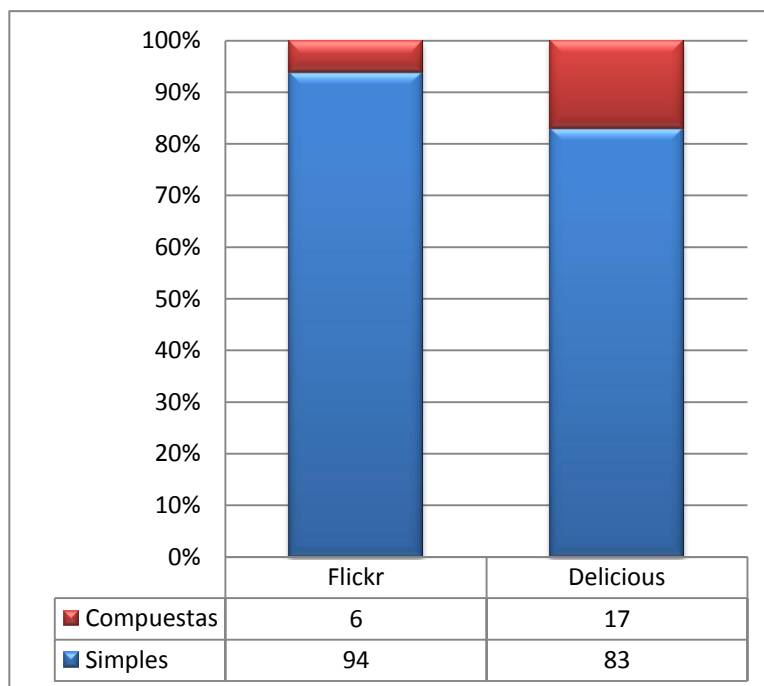


Figura 39. Valores de aparición de expresiones compuestas en Flickr y Delicious

La aparición de expresiones compuestas es un problema común en las folksonomías, ya que la falta de regularización deja al usuario sin indicaciones de cómo formar las etiquetas, lo que da lugar a distintas formas de separar las palabras, a términos muy extensos, o a conceptos equivalentes representados con las mismas palabras pero de forma diferente (cocina fácil; cocina\_facil; cocina-facil; cocinafacil).

En ambos casos el valor de aparición de expresiones compuestas es bajo, sin superar el 20% de las etiquetas. Como las expresiones compuestas son causa de una elevada dispersión del etiquetado, la aparición de valores bajos implica que no causan problemas en el proceso de recuperación de información y, aunque el objeto al que enlazan no sea devuelto, se pierde una cantidad poco significativa de información.

### 3.1. Resultados obtenidos frente a resultados anteriores

El estudio sobre los problemas que afectan a las folksonomías ha sido más teórico que práctico, e incluso entre los segundos aparecen diferentes acercamientos, muy relevantes en su campo pero de limitado uso para este trabajo en concreto.

El punto de vista de este trabajo es similar al de Guy y Tonkin, que en 2006 realizaron un trabajo centrado en Flickr y Delicious cuyo objetivo final era mejorar el funcionamiento de un sistema de tagging. Las conclusiones que obtuvieron resultan de interés en la medida en que permiten realizar una comparación con los resultados obtenidos en este trabajo, aunque en nuestro caso se desarrolla un estudio de carácter cualitativo por medio de un análisis manual.

En su estudio, Guy y Tonkin afirman que son cinco los mayores fallos en las tags de ambos portales: errores ortotipográficos; expresiones compuestas, variaciones morfológicas (conjugación de tiempos verbales, plural/singular), tags egoístas y tags de un solo uso. Ha de tenerse en cuenta que en su estudio, el análisis de las tags fue realizado por un software capaz de identificar los términos almacenados en un diccionario. Esto supuso una limitación a la hora de reconocer idiomas, expresiones populares o jergas, y etiquetas compuestas.

Entre las variaciones morfológicas, aparece la presencia de la dicotomía plural/singular; siendo el porcentaje de tags en plural en Flickr de 8%, y en Delicious 11%. Estos resultados se ajustan a los obtenidos en el presente trabajo, de 7% y 15 %.

Como ya se ha dicho, en aquel caso la identificación de idiomas estuvo sujeta a ciertas restricciones, por lo que aparece un porcentaje de tags en una lengua “desconocida” (52% en Flickr y 33% en Delicious). En este trabajo no sucede ya que todas las tags fueron identificadas.

En aquel el idioma inglés mantiene unos valores altos, de 47% en Flickr y 64% en Delicious, muy alejados de los resultados de este (97% y 94% respectivamente). Existe concordancia en cuanto a la utilización del idioma español, que continua apareciendo como el segundo más utilizado; en aquel un 3% y 2% de tags, y en este un 2,20% y 1,47%.

Entre los mayores problemas encontrados por Guy y Tonkin señalan los errores ortográficos, que afectan a un 40% de las tags en Flickr y un 28% en Delicious. En este caso no hay coincidencia con los resultados de este trabajo, ya que el valor en Flickr es nulo y en Delicious de solo el 0,92%. Sin embargo, hay una explicación para estas variaciones en base a los criterios de formación de la categoría que utilizaron. En el apartado de conclusiones señalan que han agrupado como errores ortográficos las tags con errores ortotipográficos, aquellas no identificadas por el software (idiomas o códigos) y expresiones compuestas multilingües. Por tanto, quedando explicado el origen de las diferencias, se evidencia que no hay lugar para la comparación.

Como conclusión general, los valores obtenidos ahora son consistentes con los obtenidos por Guy y Tonkin, con algunas diferencias originadas por la metodología y procesamiento de tags. El hecho de que entre ambos trabajos haya un brecha de 6 años ratifica la validez de los resultados y la teoría de que con el tiempo, se crea una suerte de patrón de comportamiento

en los usuarios, que según diversos autores (Weiss, 2005; Trant, 2009) llevará a una autonormalización en el etiquetado, aumentando la retroalimentación del sistema y mejorando la experiencia de los usuarios.

## 4. Conclusiones y futuras líneas de trabajo

Las variables estudiadas pertenecen a los que se consideran los mayores problemas lingüísticos presentes en los vocabularios no controlados. Su presencia en las muestras se daba por supuesta y, efectivamente, todas aparecen, salvo en un caso concreto. La muestra de Flickr no se ve afectada por la variable de errores ortotipográficos, lo cual sorprende, ya que incluso en el uso diario se considera normal que aparezcan este tipo de errores. El hecho de que su valor sea nulo no quiere decir que en Flickr no haya errores, aunque la imagen que se proyecta refleja la existencia de un colectivo de usuarios preocupados por la calidad y utilidad de su trabajo al describir el objeto que comparten. El valor nulo de Flickr está casi a la par del valor de Delicious, lo que hace pensar que en ambos casos la incidencia en los problemas va ser poco relevante.

El resto de las variables presentan valores similares. Ambos portales mantienen un valor medio de sinónimos y bajo de expresiones compuestas, ambas variables consideradas de gran presencia en las folksonomías. En ambos casos el valor obtenido en Flickr es inferior al de Delicious. Estos resultados evidencian que, a pesar de la presencia innegable de estos problemas, su intensidad no es tan alta como para inutilizar el proceso de recuperación de información.

El multilingüismo no es considerado uno de los mayores problemas, ya que siendo el inglés el idioma más utilizado en la web, se tiende a suponer que los usuarios sin un gran conocimiento de este adquieran un nivel suficiente para desenvolverse sin dificultad. En ambos portales el porcentaje de tags de habla inglesa supone casi el total; siendo el español el segundo idioma más utilizado. Existe coincidencia también en los cinco idiomas más utilizados, aunque hay pequeñas discrepancias en los porcentajes de los tres menos usados de cada portal. El hecho de que un mismo idioma se adjudique la mayor parte del porcentaje de tags demuestra que el nivel de multilingüismo es bajo y poco influyente en el vocabulario, por lo que no dispersa el etiquetado.

Finalmente, en la variable de la dicotomía plural/singular es donde se manifiesta una diferencia discreta entre ambos portales. La presencia de esta variable es alta en ambos, mayor en Flickr donde alcanza casi las  $\frac{3}{4}$  partes de las tags, mientras que en Delicious afecta a la mitad. La tendencia general es a usar la forma singular, más marcada sobre el total de tags que en los términos coincidentes, manteniéndose también la superioridad de Flickr. Esta diferencia solo afecta al porcentaje, que es más elevado en Flickr, pero no afecta a la tendencia a preferir la forma singular, que es característica del idioma inglés. Los valores ligeramente más elevados de Flickr pueden explicarse por el hecho de que también es mayor el porcentaje de uso del idioma inglés.

El contraste de los resultados obtenidos en este trabajo con los del estudio de Guy y Tonkin en 2006 pone de manifiesto la similitud entre ambos, a pesar de la brecha temporal. Salvando las diferencias en la metodología y procesamiento de datos, se observa que los valores de uso del idioma español y la dicotomía plural/singular son poco significativos.

A la vista de todos estos datos, se concluye que el análisis y estudio de las muestras de tags realizado permite conocer el grado con que afectan los problemas lingüísticos a la web de Flickr y Delicious. Se demuestra la presencia de estos problemas en ambas muestras, observándose que la frecuencia de aparición es baja, en general.

La forma accesible de realizar este estudio ha sido a través de muestras parciales del conjunto de tags de las folksonomías, ya que sólo los creadores o administradores de estas páginas tienen acceso al total de tags pudiendo estudiar, analizar y cuantificar la incidencia real de estos problemas. Sin embargo, las similitudes entre los resultados de muestras parciales confirman su validez, al demostrar que los valores se mantienen estables a lo largo del tiempo.

Del mismo modo, el desarrollo de estudios similares tras un periodo de tiempo programado permitiría observar nuevos cambios y variaciones en las preferencias de los usuarios a la hora de etiquetar.

Otra posibilidad de ampliar este tipo de estudio sería categorizar las tags, con el objetivo de descubrir que tipos de etiquetas se usan más. Este proceso ofrece además la oportunidad de comparar las similitudes y diferencias entre diferentes portales, teniendo en cuenta el tipo de información clasificada. Las posibilidades que ofrece este tipo de estudio han llevado a varios autores a desarrollar investigaciones respecto al tema, entre otros Enser y McGregor (1992), Jørgensen (1998), Golder y Huberman (2006), Beaudoin (2007) Abebe Rorissa (2010). Desde de un punto de vista más práctico, agrupar las etiquetas por categorías, estableciendo relaciones entre ellas e implementándolas en un portal, permitirá recuperar la información buscada y sugerir información relacionada; e incluso mostrar nubes de etiquetas agrupadas por categorías.



## Bibliografía

La redacción de la bibliografía se ha llevado a cabo según la norma ISO 690 e ISO 690-2, con una pequeña alteración: se ha incluido al inicio de cada cita la fecha utilizada durante el trabajo, para facilitar su localización.

ADELL ORTEGA, et al (2009). Efectos de la Importancia del Usuario en la Web 2.0. *UOC* [en línea]. Noviembre de 2009 [citado el 16 de enero, 2013]. Disponible en: <http://openaccess.uoc.edu/webapps/o2/handle/10609/9182> ISSN: 1696-3296

ANGELETOU, Sofia; SABOU, Marta; MOTTA, Enrico (2008). Semantically Enriching Folksonomies with FLOR. *CISWeb Workshop* [en línea], Part I: Oral Papers, pages 65-79, 2008 [citado el 13 de diciembre 2012]. Disponible en: <http://mklab.iti.gr/CISWeb/proceedings/paper5.pdf>

BEAUDOIN, Joan (2007). Flickr Image Tagging: Patterns Made Visible. *Bulletin of the American Society for Information Science and Technology*, Octubre/Noviembre 2007 [citado el 18 de diciembre de 2012]. Disponible en: <http://www.asis.org/Bulletin/Oct-07/beaudoin.html>

BERRUETA, Diego (2007). Estado del arte y propuesta de técnicas y formalismos para la anotación de contenidos y servicios por medio de folksonomías. *Morfeo-Project* [en línea]. Junio de 2007 [citado el 7 de Julio de 2012]. Disponible en: <http://forge.morfeo-project.org/wiki/images/9/96/ND.3.4.pdf>

BOIXADÓS SANUY, Joan (2012). Evaluating the Usability of a Tag-based, Multi-faceted Knowledge Organization System. *UpCommons* [en línea]. Agosto de 2012 [citado el 5 de agosto de 2012]. Disponible en: <http://upcommons.upc.edu/pfc/bitstream/2099.1/16388/1/82326.pdf>

DeveloperWorks Interviews: Tim Berners-Lee (2006). *IBM* [en línea]. Julio de 2006 [citado el 16 de enero, 2013]. Disponible en: <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206.txt>

DÍAZ PIRAQUIVE, Flor Nancy; JOYANES AGUILAR, Luis; MEDINA GARCÍA, Víctor Hugo (2009). Taxonomía, ontología y folksonomía, ¿qué son y qué beneficios u oportunidades presentan para los usuarios de la web? *Redalyc* [en línea]. Enero-junio, 2009, pp. 242-261 [citado el 25 de julio de 2012]. Disponible en: <http://redalyc.uaemex.mx/redalyc/pdf/1872/187214803010.pdf>

Folksonomía (sistemas de clasificación colectiva). *El documentalista enredado* [blog]. Octubre de 2007 [citado el 20 de octubre de 2012]. Disponible en: <http://www.documentalistaenredado.net/567/folksonomia-sistemas-de-clasificacion-colectiva/>

GARCÍA SILVA, Andrés et al (2004). Review of the state of the art: Discovering and Associating Semantics to Tags in Folksonomies. *The Knowledge Engineering Review*, vol. 00:0, 1–24, 2004 [citado el 10 de Julio de 2012]. Disponible en:

[http://oa.upm.es/6376/1/Review\\_of\\_the\\_state.pdf](http://oa.upm.es/6376/1/Review_of_the_state.pdf)

GIANNAKIDOU, Einiri (2008). Co-Clustering Tags and Social Data Sources. *The Ninth International Conference on Web-Age Information Management, 2008*, p. 317-324 [citado el 13 de diciembre de 2012]. Disponible en:

[http://researchvalue.net/repository/files/document/OSWINDS\\_2.pdf](http://researchvalue.net/repository/files/document/OSWINDS_2.pdf) ISBN: 978-0-7695-3185-4

GOLDER, S.A; HUBERMAN, B.A (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, vol. 32, núm. 2, pp. 198-208, 2006 [citado el 2 de enero, 2012]. Disponible en:

[http://pkudlib.org/gmeiCourse/files/Golder\\_usage\\_patterns\\_collaborative\\_tagging.pdf](http://pkudlib.org/gmeiCourse/files/Golder_usage_patterns_collaborative_tagging.pdf)

GÓMEZ-DÍAZ, Raquel (2012). *Etiquetar en la web social*. Barcelona, Uoc, 2012. 116 p. ISBN 978-84-9029-013-2

GUY, Marieke; TONKIN, Enma (2006). Folksonomies: Tidying up Tags? *D-Lib Magazine*, enero 2006, vol. 12, núm. 1 [citado el 7 de octubre 2012]. Disponible en:

<http://www.dlib.org/dlib/january06/guy/01guy.html>

LEONARD (2013). Diferencias entre web 1.0 y 2.0 (parte 2). *Redes Sociales* (blog). Abril de 2010. [citado el 7 de enero, 2013]. Disponible en Internet:

[http://4esoconsolacionjtorrenteymtoledo.blogspot.com.es/2010/05/diferencias-entre-web-10-y-20\\_14.html](http://4esoconsolacionjtorrenteymtoledo.blogspot.com.es/2010/05/diferencias-entre-web-10-y-20_14.html)

MATHES, Adam (2004). Folksonomies – cooperative classification and communication through shared metadata. *Adam Mathes* [blog]. Diciembre de 2004 [citado el 3 de agosto de 2012]. Disponible en:

[www.adammathes.com/academic/computer-mediated-communication/folksonomies.html](http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html)

MENGUAL ANDRÉS, Santiago (2009). T5 Nuevas tendencias: Web 2.0. *RUA* [en línea]. Diciembre de 2009 [citado el 16 de junio, 2012]. Disponible en:

<http://rua.ua.es/dspace/handle/10045/12784?mode=full&locale=es#vpreview>

MOREIRO GONZÁLEZ, José Antonio (2006). La representación y recuperación de los contenidos digitales: de los tesauros conceptuales a las folksonomías. Coordinado por Jesús Tramullas Saz. *Tendencias en documentación digital*. Gijón: Trea, 2006, p. 81-108. ISBN 84-9704-270-0

O'REILLY, Tim (2004). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *O'Reilly Media* [en línea]. Octubre de 2004 [citado el 7 de mayo de 2012]. Disponible en: <http://oreilly.com/web2/archive/what-is-web-20.html?page=1>

QUINTARELLI, Emanuele (2005). Folksonomies: power to the people. *UniMIB Meeting*, junio 24, 2005, Milán [citado el 6 de enero, 2013]. Disponible en:

<http://www.iskoi.org/doc/folksonomies.htm>

RODRÍGUEZ YUNTA, Luis (2009). Etiquetado libre frente a lenguajes documentales: aportaciones en el ámbito de Biblioteconomía y Documentación. *Ix Congreso Isko-España*, Valencia, marzo 11-13, 2009, UPV, vol 2, pp. 832-845 [citado el 16 de enero, 2013]. Disponible en: <http://digital.csic.es/handle/10261/12295?mode=full&locale=es>

RORISSA, Abebe (2010). A Comparative Study of Flickr Tags and Index Terms in a General Image Collection. *Journal of the American Society for Information Science and Technology*, 61(11), p. 2230–2242, 2010 [citado el 4 de diciembre de 2012]. Disponible en: [http://www.albany.edu/~ar955283/pubs/A\\_Comparative\\_Study\\_of\\_Flickr\\_Tags\\_and\\_Index\\_Terms\\_2010.pdf](http://www.albany.edu/~ar955283/pubs/A_Comparative_Study_of_Flickr_Tags_and_Index_Terms_2010.pdf)

SERRANO COBOS, Jorge (2006). Pasado, presente y futuro de la Web 2.0 en servicios de información digital. *BiD: textos universitaris de biblioteconomia i documentació* [en línea], diciembre, núm.17, 2006 [citado el 16 de enero, 2013]. Disponible en:

<http://www.ub.edu/bid/17serra2.htm>

SHIRKY, Clay (2004). Folksonomy. *Many2Many* [blog]. Agosto de 2004 [citado el 3 de agosto de 2012]. Disponible en:

<http://www.corante.com/many/archives/2004/08/25/folksonomy.php>

SHIRKY, Clay (2005). Ontology is overrated: categories, links, and tags. *Clay Shirky's Writings about the Internet* [blog]. Abril de 2005 [citado el 10 de julio de 2012].

Disponible en: [www.shirky.com/writings/ontology\\_overrated.html](http://www.shirky.com/writings/ontology_overrated.html)

SINHA, Rashmi (2005). A cognitive analysis of tagging. *Rashmi's Blog: Thoughts about people, technology and running a company* [blog]. Septiembre de 2005 [citado el 3 de octubre de 2012]. Disponible en: <http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/>

SPECIA, Lucia; MOTTA, Enrico (2007). *Integrating Folksonomies with the Semantic Web. Knowledge Media Institute [en línea]*, 2007 [citado el 13 de diciembre de 2012]. Disponible en:

[http://people.kmi.open.ac.uk/motta/papers/SpeciaMotta\\_ESWC-2007\\_Final.pdf](http://people.kmi.open.ac.uk/motta/papers/SpeciaMotta_ESWC-2007_Final.pdf)

TANAKA, J.W; TAYLOR, Marjorie (1991). Object categories and expertise: is the basic level in the eye of the beholder. *Cognitive Psychology* 23, 457-482, 1991 [citado el 15 de diciembre de 2012]. Disponible en:

<http://homepage.psy.utexas.edu/homepage/group/loveLAB/love/classes/concepts/TanakaTaylor1991.pdf>

THOMAS, Marliese; CAUDLE, Dana M; SCHMITZ, Cecilia M (2009). To tag or not to tag? *Library Hi Tech* Vol. 27 No. 3, 2009 pp. 411-434

TRANT, Jennifer (2009). Studying Social Tagging and Folksonomy: A Review and Framework. *Journal of Digital Information* 10 (1), 2009 [citado el 26 de agosto, 2012]. Disponible en: <http://journals.tdl.org/jodi/index.php/jodi/article/view/269/278>

VANDER WAL, Thomas (2005a). *Explaining and Showing Broad and Narrow Folksonomies*. *Off the Top* [blog]. Febrero de 2005a [citado el 7 de enero, 2013]. Disponible en: <http://www.vanderwal.net/random/entrysel.php?blog=1635>

VANDER WAL, Thomas (2005b). Folksonomy Definition and Wikipedia. *Off the Top* [blog]. Noviembre de 2005b [citado el 16 de enero, 2013]. Disponible en: <http://www.vanderwal.net/random/entrysel.php?blog=1750>

VORß, Jakob (2007). Tagging, Folksonomy & CoRenaissance of Manual Indexing? *Common Library Network GBV Göttingen, Germany* [en línea]. Enero de 2007 [citado el 25 de julio de 2012]. Disponible en: <http://arxiv.org/pdf/cs/0701072.pdf>

WEISS, Aaron (2005). The power of collective intelligence. *netWorker - Beyond file-sharing: collective intelligence* Volumen 9, número 3, 2005 [citado el 13 de diciembre de 2012]. Disponible en: <http://warburton.typepad.com/disruptive/documents/collectiveintelligenceWeiss05.pdf>

Open Directory (2012). Disponible en: <http://www.dmoz.org/>

Tesaurus de la UNESCO (2012) . Disponible en: <http://databases.unesco.org/thessp/>

Wikilengua del español (2012). Homonimia y polisemia. Disponible en: [http://www.wikilengua.org/index.php/Homonimia\\_y\\_polisemia](http://www.wikilengua.org/index.php/Homonimia_y_polisemia)

Wordnet (2012). Disponible en: <http://wordnet.princeton.edu/wordnet/>

## ANEXO

A continuación se presenta una tabla que muestra los valores totales de cada variable estudiada y del proceso de limpieza de tags.

		<i>Flickr</i>		<i>Delicious</i>		
<b>Limpieza de tags</b>	<i>Etiquetas iniciales</i>	1492		12642		
	<i>Etiquetas finales</i>	396		4825		
<b>Sinónimos</b>	<i>Muestra total</i>	296		3862		
	<i>Intralingüístico</i>	184		1686		
	<i>Interlingüístico</i>	153		2912		
<b>Plural/Singular</b>	<i>Total s/p</i>	<i>Singular</i>	2110	1040	6470	4606
		<i>Plural</i>		1070		1864
	<i>Términos coincidentes</i>	<i>Singular</i>	185	124	2886	1776
		<i>Plural</i>		61		1110
<b>Errores ortotipográficos</b>			0		91	
<b>Multilingüismo</b>	<i>Idioma más utilizado</i>	<i>Inglés</i>	1442		11885	
	<i>Principales idiomas secundarios</i>	<i>Español</i>	49	22	463	275
		<i>Francés</i>		11		77
		<i>Italiano</i>		14		62
		<i>Alemán</i>		2		49

Tabla 2. Valores totales obtenidos