



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

ETIQUETADO DE CONTENIDOS DIGITALES

Análisis y valoración de su aplicación
en el Servicio de Asuntos Sociales de
la Universidad de Salamanca

Autora: Rocío Calvo Domínguez

Tutor: Crispulo Travieso Rodríguez

Salamanca, 2015

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN
MÁSTER EN SISTEMAS DE INFORMACIÓN DIGITAL

Trabajo Fin de Máster

ETIQUETADO DE CONTENIDOS DIGITALES

Análisis y valoración de su aplicación
en el Servicio de Asuntos Sociales de
la Universidad de Salamanca

Autora: Rocío Calvo Domínguez

Tutor: Crispulo Travieso Rodríguez

Salamanca, 2015

ASIENTO CATALOGRÁFICO

TÍTULO Etiquetado de contenidos digitales: análisis y valoración de su aplicación en el Servicio de Asuntos Sociales de la Universidad de Salamanca.

AUTOR Calvo Domínguez, Rocío

DIRECTOR Travieso Rodríguez, Crispulo

PALABRAS CLAVE Recuperación de información; palabras clave; etiquetas; folksonomías; blog; Servicio de Asuntos Sociales de la Universidad de Salamanca.

KEYWORDS Information retrieval; keywords; tags; folksonomies; blog; Servicio de Asuntos Sociales de la Universidad de Salamanca.

CLASIFICACIÓN UNESCO 57 Lingüística
5701 Lingüística aplicada
570102 Documentación automatizada

FECHA 2015 - 09 – 11

RESUMEN La utilización de palabras clave en la recuperación y organización de información digital es una de las claves de la Web 2.0. Este trabajo es un estudio previo para la actualización de la folksonomía del Blog del Servicio de Asuntos Sociales de la Universidad de Salamanca, con el fin de integrar esta función en su nueva página web. Para ello se analizan las características del recurso (en concreto, la estructura por categorías y el tipo de contenidos que se publican) y evalúa tanto la lista de etiquetas anterior realizada por los profesionales de este organismo como las diversas alternativas de indización que se podrían aplicar en un futuro. El etiquetado libre, el uso de lenguajes documentales para la determinación de términos o la indización mediante sistemas automáticos son las metodologías planteadas en este estudio, que pretende determinar cuál o cuáles serían las opciones más adecuadas para describir los contenidos de la bitácora de esta institución.

ABSTRACT

The use of keywords in the recovery and organization of digital information is one of the points of Web 2.0. This paper is a preliminary study about the folksonomy of the Servicio de Asuntos Sociales de la Universidad de Salamanca's blog, for the updating of its tag list and the integration of this function on its new website. To do so, the characteristics of the resource (specifically the structure by categories and the type of content published) have been analyzed, followed by the evaluation of both the previous list of tags carried out by the professionals of this organization as the various alternatives of indexing that could be implemented in the future. Free tagging, the use of thesaurus for determining indexing terms and automatic indexing are the methodologies proposed in this paper, which aims to determine which one or ones would be most appropriate to describe the contents of the blog of this institution.

DESCRIPCIÓN

Trabajo de Fin de Máster en Sistemas de Información Digital, curso 2014-2015

SUMARIO

ASIENTO CATALOGRÁFICO.....	I
ÍNDICE DE FIGURAS.....	IV
ÍNDICE DE TABLAS	IV
INTRODUCCIÓN.....	1
1. MARCO TEÓRICO	4
1.1. El etiquetado en Internet: modalidades.	6
1.1.1. La indización automática.	6
1.1.2. El etiquetado realizado por personas.	7
2. EL SERVICIO DE ASUNTOS SOCIALES DE LA UNIVERSIDAD DE SALAMANCA	12
2.1. La página web del Servicio de Asuntos Sociales.	14
3. OBJETIVOS	17
4. METODOLOGÍA.....	18
5. ANALISIS DE CASO	22
5.1. La categorización del Blog del Servicio de Asuntos Sociales.	22
5.2. La categorización en el Blog actual	23
5.3. El etiquetado del Blog del Servicio de Asuntos Sociales.	27
5.4. Propuesta y valoración de las modalidades de etiquetado web.	31
5.4.1. Etiquetado realizado por personas en lenguaje natural	31
5.4.2. Lenguajes documentales	33
5.4.3. La indización automática	40
6. CONCLUSIONES.....	45
7. BIBLIOGRAFÍA.....	48
8. APÉNDICES.....	53
8.1. Lista de etiquetas original del Blog de Diarium del Servicio de Asuntos Sociales de la Universidad de Salamanca.	53
8.2. Muestra de los 299 unitérminos extraídos con Mallet.	60

ÍNDICE DE FIGURAS

<i>Figura 1. Esquema sobre el etiquetado en Internet (Elaboración propia).</i>	11
<i>Figura 2. Organigrama del Servicio de Asuntos Sociales de la Universidad de Salamanca.</i>	14
<i>Figura 3. Página de inicio de la antigua página web del Servicio de Asuntos Sociales.</i>	15
<i>Figura 4. Página de inicio de la actual página web del Servicio de Asuntos Sociales (versión móvil).</i>	16
<i>Figura 5. Funcionamiento del clasificador documental Mallet en el programa Símbolo de Sistema.</i>	21
<i>Figura 6. Distribución de las publicaciones por categorías en el Blog de Diarium del SAS.</i>	23
<i>Figura 7. Distribución de las publicaciones por categorías de Secciones en el Blog actual del SAS.</i>	26
<i>Figura 8. Distribución de las etiquetas por número de artículos en el Blog Diarium del SAS.</i>	28

ÍNDICE DE TABLAS

<i>Tabla 1. Distribución de las publicaciones por categorías de Áreas en el Blog actual del SAS.</i>	25
<i>Tabla 2. Número de etiquetas asociadas a nombres propios en la folksonomía creada por el SAS.</i>	30
<i>Tabla 3. Etiquetas de la folksonomía creada por el SAS presentes en los tesauros UNESCO y Eurovoc.</i>	35
<i>Tabla 4. Distribución de las etiquetas de la folksonomía del SAS en microtesauros UNESCO y Eurovoc.</i>	38
<i>Tabla 5. Términos más frecuentes del Blog del Servicio de Asuntos Sociales (resultados de Mallet).</i>	41

INTRODUCCIÓN

El presente Trabajo Fin de Máster (TFM) nace como resultado de mi estrecha colaboración con el Servicio de Asuntos Sociales (SAS) de la Universidad de Salamanca. Desde el mes de febrero del 2015, he colaborado junto con la responsable de la unidad de Participación y Voluntariado, Begoña Lázaro Blanco, en la promoción del programa Servicio Voluntario Europeo. Con posterioridad, ella sería la tutora del Prácticum del Máster de Sistemas de Información Digital, realizado en este mismo centro durante 5 semanas.

Es importante destacar que las funciones realizadas durante el período de prácticas poco tienen que ver con el objetivo de este estudio: a lo largo de las 5 semanas, mi trabajo se centró en la dinamización y reorganización de las redes sociales y recursos digitales gestionados por el SAS, independientes todos ellos de la página web, objeto de este estudio. Por ello, sólo se ha recuperado de la Memoria de Prácticum lo referente a la estructura y funciones del Servicio de Asuntos Sociales (punto 2 de este TFM).

Sin embargo, mi colaboración con el centro tanto antes como durante el Prácticum facilitó notablemente la elaboración de este estudio, ya que sin la participación e interés de los profesionales del Servicio de Asuntos Sociales y muy concretamente de la tutora Begoña Lázaro sería imposible haber realizado este trabajo. Espero, en agradecimiento, que el presente análisis pueda ser de utilidad para el SAS.

- **Justificación**

El personal de esta institución se conforma fundamentalmente por especialistas en diversos ámbitos sociales. La ausencia de profesionales de la Documentación o de un personal formado en este campo ha llevado a que el Servicio de Asuntos Sociales cuente con notables carencias en lo que respecta a la organización y gestión de los contenidos digitales.

Uno de los problemas planteados es el hecho de que en la nueva página web del SAS, inaugurada en el mes de abril, y más concretamente en su bitácora (importada del antiguo Blog de la plataforma de la Universidad de Salamanca Diarium) no se utilizan palabras para la recuperación de la información. El motivo alegado es que, pese a contar con una lista de etiquetas anterior y estar importada al nuevo recurso, esta no se considera adecuada.

Pese a que la ausencia de esta funcionalidad es percibida por el personal como un problema, los profesionales no han podido realizar una evaluación detenida de la anterior lista y de los contenidos de su blog, por carecer fundamentalmente de tiempo. Es en este contexto en el cual surgió la idea de este Trabajo Fin de Máster, en el que se ha procurado, con su ayuda, analizar en detalle estos aspectos.

- **Objetivos**

Así, el objetivo principal de este Trabajo Fin de Máster nace de una meta marcada por el Servicio de Asuntos Sociales, que sería la eventual creación de una nueva lista de etiquetas, completa, actualizada y profesional, para su inclusión dentro del blog de la institución.

Para ello, se deben plantear una serie de objetivos específicos. Resulta esencial la realización de un estudio previo que analice las características del Blog del Servicio de Asuntos Sociales, lo que implica:

- Analizar la estructura por categorías existente en el Blog antiguo y en el actual.
- Evaluar la antigua lista de etiquetas importada del blog de Diarium, para comprobar el grado de adecuación al recurso.

Tras este análisis, cabría:

- Estudiar las diferentes alternativas a la hora de realizar una indización de los contenidos del Blog actual, ajustándose a los requerimientos de la información y las características del Servicio de Asuntos Sociales.
- Evaluar las diversas opciones en este caso concreto, comparando las conclusiones con lo dictaminado a nivel teórico por los especialistas.

- **Metodología**

Tras una revisión bibliográfica, base del Marco Teórico de este TFM en el cual se han expuesto las diferentes modalidades de etiquetado que se utilizan a día de hoy en Internet, se ha estructurado el análisis. Partiendo de una descripción del sistema de categorías en el Blog antiguo y actual y la exposición de una serie de conclusiones sobre dicha organización, el TFM se centra en el estudio del etiquetado. A este respecto, se han descrito las características, uso, pros y contras de la lista de etiquetas importada de Diarium para posteriormente, analizar tres posibles alternativas:

- La indización basada en el lenguaje natural, que a su vez puede ser personal, grupal o social. Se tienen en cuenta para ello lo expuesto por la teoría, la percepción de los profesionales del Servicio de Asuntos Sociales y los problemas planteados en la antigua lista de etiquetas.
- La indización basada en el lenguaje documental, mediante la comparativa de la antigua lista de etiquetas y su presencia en los tesauros de la UNESCO y Eurovoc, siguiendo la metodología planteada por Rodríguez Yunta (2009).
- La indización semiautomatizada, mediante el análisis de los 299 unitérminos más frecuentes dentro del conjunto del Blog del SAS, según los resultados del clasificador documental Mallet.

A modo de conclusión, se han analizado las ventajas y limitaciones de estas modalidades en su conjunto, comparando lo planteado a nivel teórico con la realidad del recurso y así evaluar cuál podría ser la solución más recomendable a la hora de crear una lista de etiquetas para el Blog del Servicio de Asuntos Sociales.

- **Estructura**

El presente Trabajo Fin de Máster consta de los siguientes apartados:

- La presente **Introducción**, en la que se expone la justificación, al mismo tiempo que se avanzan los objetivos, la metodología y la estructura del trabajo.
- **Marco teórico**, centrado en la recuperación de información en Internet y el peso de las palabras clave en esta labor, junto con la relación de las diversas modalidades de etiquetado.
- **El Servicio de Asuntos Sociales de la Universidad de Salamanca**, en el que se realiza una presentación del SAS y las funciones que desempeña la institución, junto con una exposición sobre la actual página web del Servicio.
- **Objetivos**, donde se expone el objetivo principal y los específicos de este Trabajo Fin de Máster.
- **Metodología**, con una descripción y justificación detallada de los pasos seguidos tanto en la revisión bibliográfica como en el estudio de caso.
- **Estudio de caso**, centrado en el análisis de la organización por categorías, el uso de palabras clave en el antiguo Blog del Servicio de Asuntos Sociales y el estudio de las modalidades aplicables a la hora de indizar el nuevo recurso.
- **Conclusiones**, donde se recogerá el análisis final de los resultados obtenidos en función de los objetivos planteados en un principio.
- **Apéndices**, con una relación de la lista de etiquetas previa del Servicio de Asuntos Sociales y otra con los 299 unitérminos extraídos del actual Blog del SAS gracias al clasificador documental Mallet.

1. MARCO TEÓRICO

No es complicado encontrar textos de los años 90 en los que profesionales de la Documentación describen Internet como *even worse than a vandalized library* (Michael Gorman, 1995). Ese mismo año, el economista francés Jacques Attali escribe en el periódico *Le Monde* un artículo en defensa del laberinto cibernético, ya que considera que la complejidad de Internet no deja de ser un ejemplo de la diversidad informativa que la sociedad tiene por naturaleza; es más, considera que la creación de *lignes droites* rompe con la armonía, con la belleza bruta de Internet.

Admiremos o no el caos cibernético, lo cierto es que igualmente por naturaleza, el ser humano necesita el orden, o al menos, encontrar vías que le faciliten obtener lo que desea con el mínimo esfuerzo, optimizando tiempo y recursos. Desde la Antigüedad, por ejemplo, las bibliotecas se vieron obligadas a la creación de inventarios y catálogos para identificar y localizar sus fondos, una tarea que a lo largo del tiempo se fue perfeccionando hasta lograr las conocidas *Reglas de Catalogación*, dotando de estructura a lo que hasta entonces era una mera enumeración. Asimismo, era necesario conocer el contenido de los libros: esa función la realizaban los resúmenes de las obras, cuya relevancia se mantiene hasta nuestros días. Otra metodología sería la indización, o lo que es lo mismo, la selección de conceptos que nos permitan describir el contenido de los textos, “bien procedentes del lenguaje natural empleado por los autores, o de un lenguaje documental previamente seleccionado” (María Pinto, 1993), como pueden ser los tesauros, los encabezamientos de materias, etc.

Con la aparición de Internet se abre un nuevo paradigma ante el cual volvemos al punto de partida y se deben encontrar soluciones ante el crecimiento exponencial y caótico de la información digital. En un primer momento, los profesionales de la Documentación, que hasta entonces habían sido los responsables últimos de la organización de los contenidos, se consideraron esenciales en esa labor. En el texto citado anteriormente de Gorman, su conclusión final es que *Librarians and, especially, catalogers are uniquely qualified to tame the electronic wilderness. If not us, who? If not now, when?* Efectivamente, se trataba de una necesidad inmediata, lo que explica que en esta misma época surjan los sistemas de búsqueda de información digital; muchos de ellos directorios web, como era el caso de Yahoo, en los cuales los enlaces son seleccionados, valorados y categorizados de manera manual y por tanto, el papel de los profesionales de la Documentación podría ser relevante.

Sin embargo, la aparición de los buscadores (a destacar el caso de Google) rompen con toda esa lógica: expertos en computación crean un algoritmo (en este caso el *PageRank*) mediante el cual el usuario introduce una serie de palabras clave y con ello obtendrá resultados relacionados con su búsqueda que son detectados de forma automática dentro del conjunto superficial de la Red, ya que muchos contenidos en Internet forman parte de la conocida como *Deep Web*, con recursos a los que no podemos acceder libremente mediante un buscador.

Se podría pensar que con la aparición de Google y buscadores similares, con sus inteligentes *crawlers*, y la instalación de sistemas de búsqueda dentro de los sitios web ya no tendríamos que preocuparnos por el caos cibernético, y por tanto, la categorización e indización de la información web sería innecesaria. Es más, no es raro encontrar artículos en Internet en los que expertos en mejorar el posicionamiento de enlaces dentro de los buscadores (lo que se conoce como el SEO, *Search Engine Optimization*) consideran que el hecho de que una página web describa su contenido mediante palabras clave no mejorará notablemente su visibilidad¹, pudiendo ser incluso contraproducente, si por ejemplo, utilizamos demasiadas veces conceptos similares.

Google deja claro que un uso excesivo o ineficaz de los descriptores en un sitio web puede llevar a que el buscador considere los contenidos como spam. Por tanto, la clave radica en realizarlo de manera correcta, lo que para esta compañía consiste en describir tal y como busca el usuario: al fin y al cabo, la información suele buscarse mediante conceptos clave y no tanto por el contenido. No cabe olvidar que esta empresa cuenta con un servicio de pago denominado “Planificador de Palabras Clave” dentro de su plataforma Google Adwords, que nos permite buscar descriptores óptimos en función de las búsquedas más realizadas, lo que resulta esencial, por ejemplo, en las campañas de marketing: las empresas pujan para estar en las primeras posiciones ante una determinada búsqueda, y por tanto, el orden de los resultados no está únicamente determinado por la relevancia.

De todos modos, reducir el papel de la descripción en Internet a la influencia que esto tendría en los buscadores sería limitar notablemente su función. Hasta los más escépticos con las palabras clave reconocen su utilidad dentro de las propias páginas web: nos permite la recuperación de información sobre una determinada temática o incluso tener secciones específicas dentro de categorías genéricas. La indización de artículos mejora notablemente la imagen del recurso a nivel profesional, ofreciendo al usuario un servicio a mayores (la búsqueda mediante palabras clave). Por último, los recursos pueden contar con listas o nubes de etiquetas públicas, que permiten observar de manera automática las temáticas frecuentes tratadas en la página web.

La clave, por tanto, estaría en cómo y quién indiza la información digital. De nuevo, esa responsabilidad no caerá exclusivamente en manos del documentalista, debido principalmente a la diversidad temática de contenidos y a la creación exponencial de información en Internet. Ante la necesidad de describir ciertos contenidos, autores y/o usuarios empiezan a ocuparse de esta tarea; así, se pasa de la indización profesional al etiquetado, carente de estructura pero adaptado a las múltiples y específicas temáticas que podemos encontrar en Internet.

¹ Artículos como los citados en la bibliografía de los blogs *Técnica Seo* (2014) y *SEM, SEO y mucho +* (2015) nos remiten a las declaraciones realizadas en 2009 por el director del departamento contra el spam en web de Google, Matt Cutts (Disponible en: <https://www.youtube.com/watch?v=jK7IPbnmvVU#t=82>). En este vídeo, afirma: *We don't use the keywords meta tag in our search ranking. Other search engines might, but Google doesn't. (...) Too many people have spammed that too much. We really just don't use this information at all.*

Esto no sólo incluye páginas web: las redes sociales representan el paradigma de la implementación de etiquetas temáticas. Cabe destacar el caso de Twitter o Instagram, donde las publicaciones pueden incluir hashtags, lo que no deja de ser una etiqueta, y a través del buscador del propio sitio podemos encontrar las publicaciones procedentes de perfiles públicos que están utilizando esa palabra. Igualmente, Twitter nos ofrece la posibilidad de consultar lo denominado *como Trending Topic*, o sea, las etiquetas más utilizadas en tiempo real en un determinado lugar. Esto lleva a que el hashtag pueda ser utilizado de forma viral con todo tipo de finalidades: la meramente informativa, la publicidad o la propaganda política, entre otros usos.

1.1. El etiquetado en Internet: modalidades

A la hora de hablar del etiquetado en Internet, encontramos diferentes métodos en función de quién y cómo este se realiza.

En función del quién lo realiza, se debe diferenciar en un primer nivel, como indica Hassan Montero (2006), dos grandes categorías: la **indización humana** y la **indización automática**. En la primera, una persona o grupo de personas determina los conceptos más adecuados para describir los contenidos, mientras que en el segundo caso, esto viene determinado por algoritmos que calculan el peso de cada palabra en un documento o serie de documentos.

Dentro de la indización humana se diferencia entre el lenguaje controlado y el lenguaje natural. Así, tendríamos los tesauros, realizados por profesionales de la Documentación, que pueden ser eventualmente utilizados para la asignación de palabras clave en Internet, y el etiquetado realizado por el autor/es y/o usuarios de la información, que designan libremente los términos que consideran más apropiados.

1.1.1. La indización automática

De entrada, cabría analizar las ventajas e inconvenientes de cada una de estas metodologías. La automatización del proceso de etiquetado tiene su base en algo ya expuesto: el aumento exponencial de la información en la Red. El coste de tiempo y dinero que supondría dejar exclusivamente en manos de personas (profesionales o no de la información) esta tarea sería inasumible, lo que ha llevado a que la información se “semi-automatice”. Un ejemplo de esto sería lo que se produce al subir una fotografía en Facebook. El recurso, automáticamente, nos da la opción de etiquetar a personas en esa imagen; es más, puede que incluso detecte a la persona. Este sistema facilita la tarea y reduce considerablemente la carga de trabajo por parte del usuario, pero al fin y al cabo, es la persona la que decide si incluye o no esa etiqueta; de hecho, si el recurso es capaz en un momento dado de identificar a amigos, se debe al trabajo previo y/o de corrección por parte del usuario.

Dejar la responsabilidad del etiquetado web en manos de un sistema automático tendría evidentes consecuencias negativas. Por ejemplo: a la hora de trabajar con textos, detectaríamos como palabras más relevantes lo que se conoce como “palabras vacías”, ya que términos como artículos, preposiciones o pronombres son los más

recurrentes; sin embargo, estos carecen de valor descriptivo. Para evitar esto, los programas suelen incluir ficheros con palabras a evitar; son conceptos generalmente seleccionados por personas, otra muestra más de la importancia de la semi-automatización y de “la subjetividad de la interpretación del contenido que pueden dar los distintos profesionales” (Gómez Díaz, 2012). Igualmente, detecta exclusivamente las palabras escritas por el autor de la información, cuando con una correcta revisión se podrían encontrar términos genéricos más adecuados a la hora de ser utilizados como etiquetas.

De todos modos, hay evidentes problemas en el tratamiento semi-automático del etiquetado en Internet. Para empezar, estos sistemas no detectan palabras sino “cadenas de caracteres” (Moreiro González, 2007). Como nos indica este autor, no se detecta la palabra gato, sino g-a-t-o, lo que lleva a que los programas no sean capaces a apreciar la complejidad del lenguaje natural: la traducción de términos (chat, cat...), la sinonimia (minino), la polisemia (ya que puede hacer referencia a la máquina para elevar pesos) o las palabras derivadas (gatuno), entre otros ejemplos. Otro problema presentado por estos sistemas automatizados reside en el hecho de que sólo detecta unitérminos, lo que lleva a que conceptos que van generalmente asociados no sean detectados como una sola palabra clave (ej. Oxfam Intermón). Por último, el español como lengua tiene, caracteres especiales como son las tildes, la diéresis o letras como la ñ, por lo que en sistemas diseñados en lengua inglesa los términos no son siempre detectados.

1.1.2. El etiquetado realizado por personas

En cuanto al etiquetado realizado por personas, debemos hacer una diferencia esencial entre el realizado a partir de un lenguaje controlado y el etiquetado en lenguaje libre o natural, mucho más habitual en el entorno web.

▪ El lenguaje controlado

La utilización de un lenguaje controlado para la descripción conceptual de una página web tendría, según Frederick Lancaster (2002) dos claras ventajas: “facilitar la representación consistente de las materias (...) evitando la dispersión de los elementos relacionados” (lo que lleva a la agrupación de sinónimos, la diferenciación de las palabras homónimas, etc.) y “facilitar la realización de una búsqueda amplia sobre una materia enlazando los términos con relaciones paradigmáticas o sintagmáticas”. El tesoro tiene, sin duda, la ventaja de que ante el etiquetado sobre una determinada temática, podamos contar con una lista predeterminada de conceptos, e igualmente, si el usuario tiene acceso a esa lista “le será más fácil identificar los términos apropiados a una necesidad de información”.

Sin embargo, el propio Lancaster apunta como principal problema que los lenguajes controlados no permiten la especificación terminológica (concretamente habla de la utilización de nombres de ciudades, pero se aplica igualmente a cualquier nombre propio), por lo que las etiquetas a utilizar podrían ser muy genéricas para describir la información digital. La problemática, sin embargo, va mucho más allá: a día de hoy los tesauros no representan la diversidad informativa que genera Internet y la creación de

lenguajes controlados de cada temática es inasumible para los profesionales de la Documentación, entre otras motivos porque obliga a un exigente trabajo intelectual que requiere tiempo; teniendo en cuenta la velocidad con la que se generan conceptos en la Red, es inasumible que el lenguaje controlado esté debidamente actualizado.

Con respecto a esto, hay un consenso generalizado por parte de los autores en cuanto a que los tesauros a día de hoy están obsoletos, ya que al margen del etiquetado en Internet, numerosos términos genéricos de uso común en nuestros días siguen sin estar incluidos en estos lenguajes controlados. Por último, cabría señalar lo apuntado por Gómez Díaz (2012): “puede haber distintos tesauros para una misma área temática que recojan términos diferentes y por lo tanto un mismo documento (...) dependiendo del enfoque del tesoro, la misma realidad se representará con palabras distintas”.

▪ El lenguaje natural

El etiquetado que nace del lenguaje natural aglutina toda la indización no realizada por profesionales de la Información: por tanto, podemos hablar aquí tanto del etiquetado realizado por autores como por usuarios, e igualmente, del etiquetado personal al social, pasando por el grupal; esto lleva a que debemos incluir dentro de esta categoría el origen de prácticamente todo el etiquetado que se realiza a día de hoy en Internet.

Las ventajas e inconvenientes de la utilización del lenguaje natural en la Web son inversamente proporcionales a las señaladas en el caso de los tesauros. Así, entre las ventajas tenemos el uso ilimitado de términos, lo que permite la especificación en la descripción; la responsabilidad en cuanto a la determinación de los términos se socializa, junto con el hecho de que los conceptos están perfectamente actualizados, ya que es el uso que se haga de los conceptos lo que determina si es un término actual u obsoleto. Sin embargo, se carece del orden y el control que aportan los tesauros, por lo que un problema habitual es el uso de redundancias en las etiquetas o de ambigüedades (por ejemplo, debido a homonimias), por no hablar de otros errores frecuentes como las erratas, las faltas de ortografía, el uso de todo tipo de variantes de una palabra, etc. Igualmente, la ausencia de una estructura temática lleva a que cualquier término pueda ser utilizado en cualquier tipo de contexto, sea adecuado o no, unido a que el usuario no puede recurrir a una lista conceptual que sirva de referencia a la hora de saber cómo recuperar la información. De alguna manera esta carencia justifica que en las páginas webs existan las nubes de etiquetas; por tanto, no cabe desdeñar el papel del etiquetado como parte esencial de la recuperación de información.

Dentro de la indización basada en el lenguaje natural, podemos diferenciar diferentes tipos en función de las personas involucradas en el proceso. Por ejemplo, si lo hacemos por el número de personas, podemos hablar de:

- Etiquetado personal (realizado por una única persona)
- Etiquetado “grupal”: el realizado por un determinado grupo de personas
- Etiquetado social o libre, también denominado como Indización Social Agregada o Etiquetado Colaborativo, en el cual cualquier persona podría contribuir a esta tarea.

Javier Cañada (2006)² apunta, a su vez, a 4 tipos diferentes de etiquetado basado en lenguaje natural, en función de la metodología y/o el objetivo perseguido:

- El etiquetado egoísta: aquel en el que el usuario etiqueta para la gestión personal del contenido y no tanto para uso social, sea con términos concretos o genéricos. Un ejemplo sería la etiqueta “Sin revisar”: tiene sentido para el autor, pero no necesariamente para el resto de la comunidad.
- El etiquetado amiguista: en este, el autor de la etiqueta debe pensar a la hora de elegir el término en un número determinado de personas, unidas por un interés común. Un ejemplo sería “Fiesta Graduación 2014”. Este tipo de etiquetado tiene una orientación social limitada, pero “permite compartir y reforzar la pertenencia entre el grupo” (Pérez Sanchidrián, 2014).
- Etiquetado altruista: cuando el autor de la etiqueta tiene el objetivo de compartir la información con los usuarios, lo que lleva a que utilice términos genéricos y verdaderamente descriptivos. Sin duda, sería la metodología más recomendable, pero es la menos habitual. Se utiliza, por ejemplo, en los diarios en versión digital para indizar las noticias, con etiquetas como “Justicia”, “Real Madrid”, etc.
- Etiquetado populista: aquel en el que el objetivo es que el contenido se recupere de forma masiva, por lo que se utilizan etiquetas subjetivas y poco descriptivas, pero ampliamente utilizadas en Internet. Este tipo de etiquetado es muy habitual en redes sociales: por ejemplo, el uso de hashtags en Instagram como #happy o #lovely en todo tipo de fotografías garantiza que puedan ser visualizadas por un número potencialmente alto de personas.

El resultado final de este trabajo de indización humana se conoce con el nombre de **folksonomía**, término acuñado por Thomas Van der Wal, gracias a la conjunción de la palabra folk (gente, pueblo) y taxonomy (taxonomía, clasificación). Este autor diferencia dos tipos de folksonomías: la amplia (*Broad Folksonomy*), en la que el proceso de etiquetado se realiza de manera abierta y por tanto los lectores tienen poder a la hora de indizar la información, y la estrecha (*Narrow Folksonomy*), donde los autores, o en todo caso, un grupo reducido de personas, tienen la responsabilidad del etiquetado. La creación de folksonomías, es, sin lugar a dudas, uno de los elementos clave que conforman la Web 2.0, o lo que es lo mismo, el Internet actual en la cual el usuario asume un rol activo a la hora de generar contenido.

² Citado en Gómez Díaz (2012). Fuente original: Tipologías y estilos en el etiquetado social en *Terremoto.net: Diseño de interacción desde el año 2000* no disponible en la actualidad. (Link: <http://www.terremoto.net/tipologias-y-estilos-en-el-etiquetado-social/>).

Como se puede comprobar, ninguna técnica utilizada a día de hoy en el etiquetado en Internet se considera perfecta, ya que todas las posibilidades tienen pros y contras. Teniendo en cuenta esto, lo más recomendable a la hora de decidir qué modelo seguir para etiquetar un determinado recurso podría ser la realización de un estudio previo que nos permita determinar cuáles son las técnicas más adecuadas en función del recurso, pudiendo, en la medida de lo posible, combinar varios métodos. Es así cómo nace la idea de este Trabajo Fin de Máster, que intenta analizar y si es posible dar una solución a un problema real: el etiquetado en el blog integrado en la página web del Servicio de Asuntos Sociales de la Universidad de Salamanca.

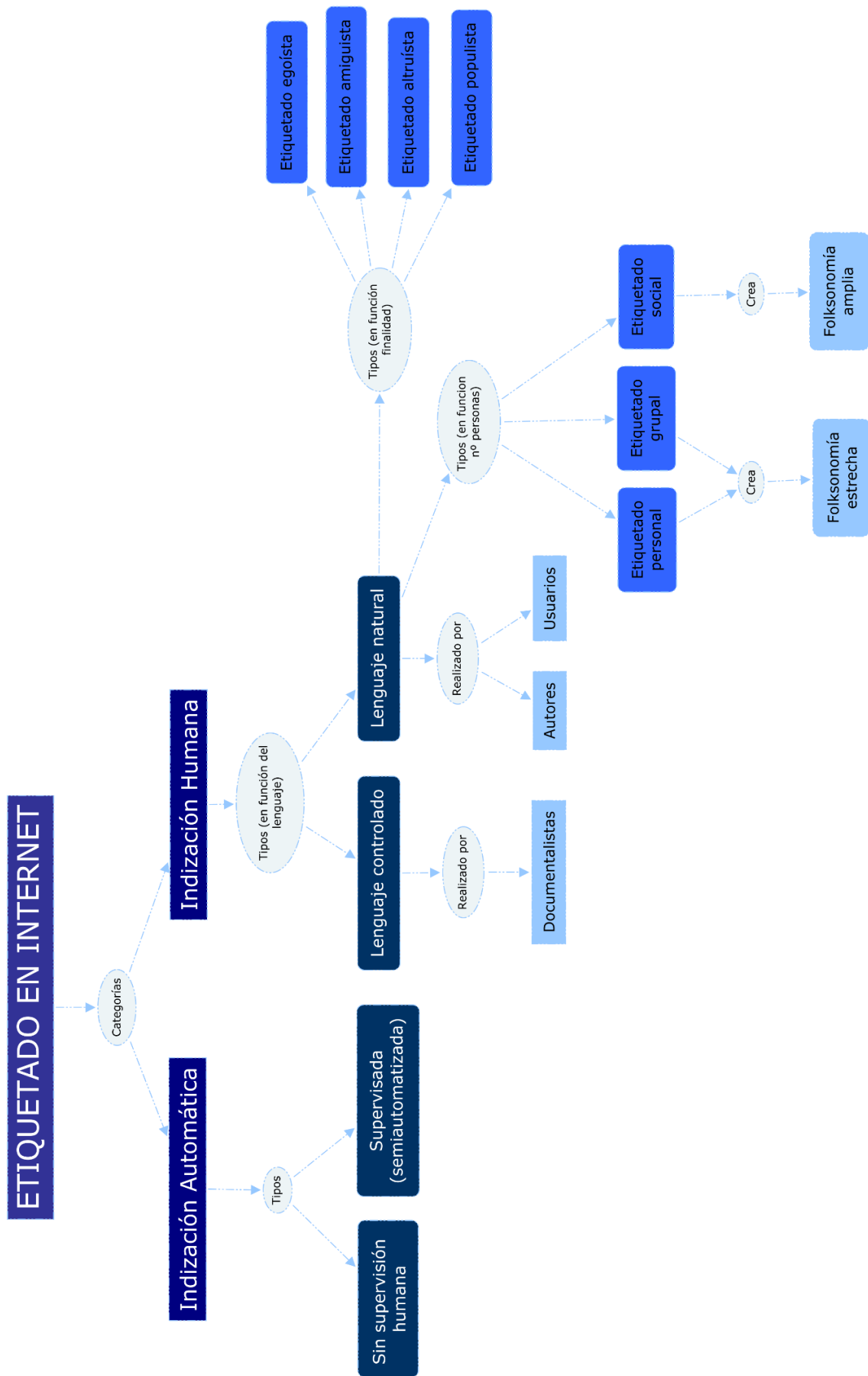


Figura 1. Esquema sobre el etiquetado en Internet (Elaboración propia).

2. EL SERVICIO DE ASUNTOS SOCIALES DE LA UNIVERSIDAD DE SALAMANCA

El Servicio de Asuntos Sociales (SAS) de la Universidad de Salamanca es una unidad dependiente del Vicerrectorado de Atención al Estudiante y Extensión Universitaria. Su labor se centra en dos ámbitos: el asesoramiento y apoyo que el conjunto de la comunidad universitaria pueda requerir para contribuir a su mejor integración, participación y bienestar personal y social durante su vida universitaria, y la promoción del compromiso social y responsable con la sociedad y la defensa de los derechos humanos a través de acciones de voluntariado que se pueden realizar en el entorno local y/o en otros países.

Dentro del organigrama del SAS podemos encontrar diversas unidades especializadas en diversos ámbitos de acción:

- **Apoyo Social y Atención a Estudiantes Extranjeros.** Desde aquí se realizan intervenciones ante problemas sociales en los que la comunidad universitaria necesite asesoramiento o una intervención técnica, al mismo tiempo que se ayuda a todos aquellos estudiantes extranjeros con problemas tales como la estancia legal en el país, cuestiones médicas, etc. Igualmente, esta unidad suele organizar campañas de difusión sobre la donación de sangre o médula (por ejemplo, el Día del Donante Universitario).
 - **Programa Intergeneracional.** Dentro de Apoyo Social, aunque funciona de manera semiindependiente, esta unidad se centra en dos actividades: el alojamiento de jóvenes universitarios en las casas de personas mayores (a través del Programa de Alojamientos Compartidos) y la realización de actividades que promueven el intercambio intergeneracional de experiencias (talleres de risoterapia, cursos de inglés, plantas medicinales, juegos...).
- **Apoyo a la Comunidad Universitaria con Discapacidad.** Tiene como objetivo garantizar el principio de igualdad de oportunidades, facilitando el acceso a la enseñanza universitaria de personas con discapacidad, adaptando los currículos a su situación particular, etc. Una vez terminados los estudios universitarios, igualmente se ofrece a los jóvenes un servicio de mediación que facilite su integración en el mercado laboral. Esta unidad dispone de intérpretes en lenguaje de signos que pueden ayudar a estudiantes con discapacidad auditiva y que asisten a todas aquellas actividades donde se considere necesaria su presencia. Por último, los responsables de este servicio realizan diversas campañas como por ejemplo la del Día de la Discapacidad.
- **Voluntariado y Participación.** El objetivo principal de esta unidad consiste en lograr la colaboración de la comunidad universitaria en iniciativas sociales promovidas por entidades no lucrativas de Salamanca, Zamora y Ávila con las que se han firmado acuerdos de colaboración, una experiencia que puede ser posteriormente homologada mediante un certificado. Igualmente, el servicio realiza talleres de formación básica en voluntariado; participa en el programa de Erasmus+ Servicio Voluntario Europeo como organización de envío, coordinadora y de acogida de voluntarios; y organiza campañas de promoción

de las acciones voluntarias (Ferias de Bienvenida, Día Internacional del Voluntariado, jornadas de las Redes de Voluntariado de Salamanca y Zamora...). Por último, señalar que la responsable de esta Unidad (Begoña Lázaro) gestiona en solitario las redes sociales del Servicio de Asuntos Sociales (Facebook, Twitter, Flickr, Google+, Youtube), al igual que el programa de Radio USAL "Dejando Huella".

- **Atención Psicológica y Salud Mental.** Se centra en los aspectos relacionados con la salud mental, el aprendizaje y la adaptación al entorno académico y profesional, pero también la atención psicológica y de prevención ante problemas personales y/o relacionales que puedan manifestarse con especial gravedad, especialmente en los más jóvenes (por ejemplo, ante problemas relativos a la sexualidad, la vida en pareja, la sociabilidad...).

 - **Dificultades del lenguaje oral y escrito.** Se trata de un servicio de evaluación e intervención dirigido a alumnos de infantil, primaria y secundaria además de a universitarios que tengan algún tipo de problema de este tipo, fundamentalmente de la provincia de Ávila (el servicio se presta exclusivamente en la Escuela de Educación y Turismo de esta ciudad).

- **Psiquiatría, Psicología Médica y Salud mental.** Esta unidad está centrada en la atención de aquellos casos graves que requieren una intervención psiquiátrica, sea por problemáticas mentales, emocionales o sexuales.

 - **Programa USALudable.** Tiene como finalidad la promoción de la salud y la prevención de riesgos psicosociales en la comunidad universitaria (drogas y otras adicciones, la sexualidad, la violencia en las relaciones interpersonales y la nutrición). Para ello se realizan campañas de sensibilización, además de contar con un servicio de atención personalizada online y presencial para cualquier información o duda sobre tratamientos especializados.

- **Oficina de Cooperación.** Se encarga de los proyectos y actividades que fomentan la Cooperación al Desarrollo tanto dentro de la Universidad (mediante campañas solidarias como la del Euro Solidario a la hora de realizar la matrícula en la Universidad) como en el exterior con programas que permiten la colaboración en los países menos desarrollados de Asia, África o América (Becas Manuel Andrés Sánchez para los jóvenes universitarios y ayudas para la realización de Acciones de Cooperación Universitaria para el Desarrollo destinada a todos los miembros de la Universidad de Salamanca). Además, desde aquí se gestionan proyectos pertenecientes a otras instituciones, como es el caso del Voluntariado Universitario de las Naciones Unidas y de la Agencia de Cooperación Internacional para el Desarrollo (AECID). Por último, se realizan jornadas de sensibilización sobre las problemáticas de países en vías de desarrollo.

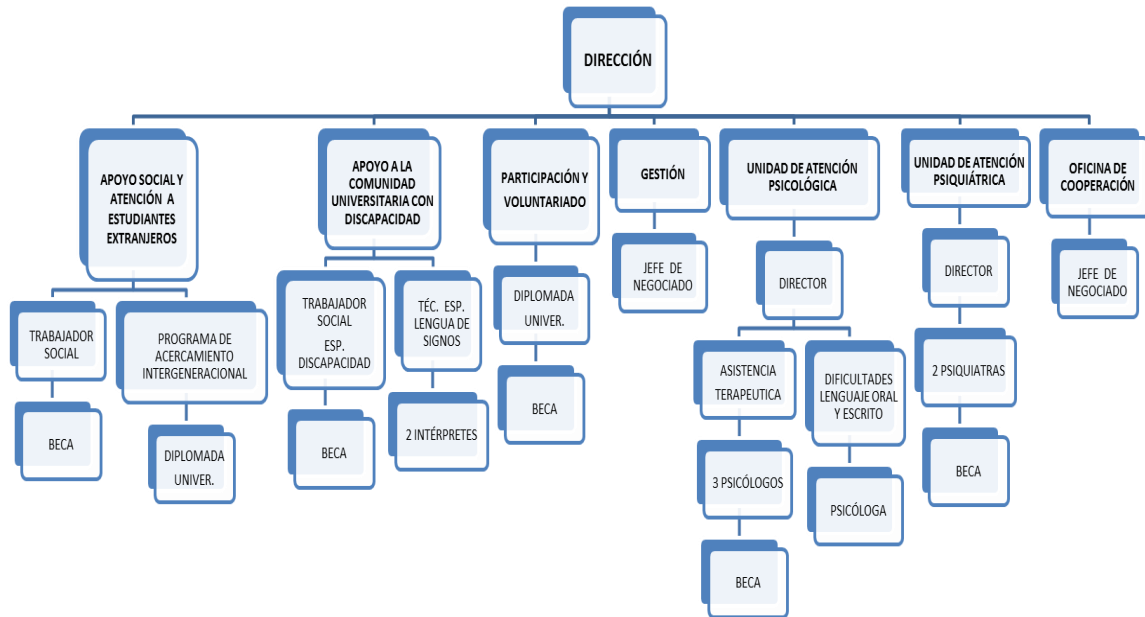


Figura 2. Organigrama del Servicio de Asuntos Sociales de la Universidad de Salamanca.

Cabe señalar, por último, que el Servicio de Asuntos Sociales, como conjunto, participa en diversas redes de trabajo:

- A nivel estatal, en diversos grupos de trabajo de la Conferencia de Rectores de las Universidades Españolas (CRUE), como es el caso de la Red de Servicios de Apoyo a Personas con Discapacidad en la Universidad, la Red de Universidades Saludables y el Observatorio de Cooperación Universitaria al Desarrollo.
- A nivel autonómico, forma parte del Programa de Acercamiento Intergeneracional en colaboración con la Junta de Castilla y León y los ayuntamientos de Ávila, Zamora y Salamanca.
- A nivel municipal, en Salamanca, en las Mesas de Voluntariado y de Prevención de la Drogodependencia.

2.1. La página web del Servicio de Asuntos Sociales

El Servicio de Asuntos Sociales cuenta con una nueva página web desde el mes de abril del presente año. Hasta esta fecha, el sitio utilizaba una plantilla del sistema de gestión de contenidos (CMS - *Content Management System*) Joomla, en la cual se podía encontrar la información general de los servicios prestados en esta institución. De forma paralela, un blog en la plataforma Diarium informaba de las novedades del Servicio; de todos modos, algunas de las novedades se publicaban igualmente dentro de la página principal del SAS (especialmente, las relativas a la unidad de Participación y Voluntariado).

Con el fin de aunar la información en un mismo recurso se optó por el cambio de plataforma, optando en este caso por el CMS Wordpress, ya que a diferencia de Joomla, este permite no sólo la integración de la página web y el blog, sino también que la información de la bitácora pueda ser visualizada desde la página de inicio del sitio (en la sección Destacados). Otro problema planteado por el antiguo formato era la ausencia de tecnología Responsive, que permite la adaptación de la plantilla web a los dispositivos móviles (tanto tablets como smartphones). A día de hoy, la utilización de plantillas basadas en el Responsive Web Design resulta esencial por el uso creciente de este tipo de aparatos, además de que permite la creación de contenidos y su adaptación automática al formato de pantalla, lo que favorece su visualización.

Los profesionales del Servicio han señalado igualmente que a las cuestiones técnicas hay que añadir que el cambio de imagen corporativa experimentado en el último año obligaba igualmente a una modernización del sitio web.

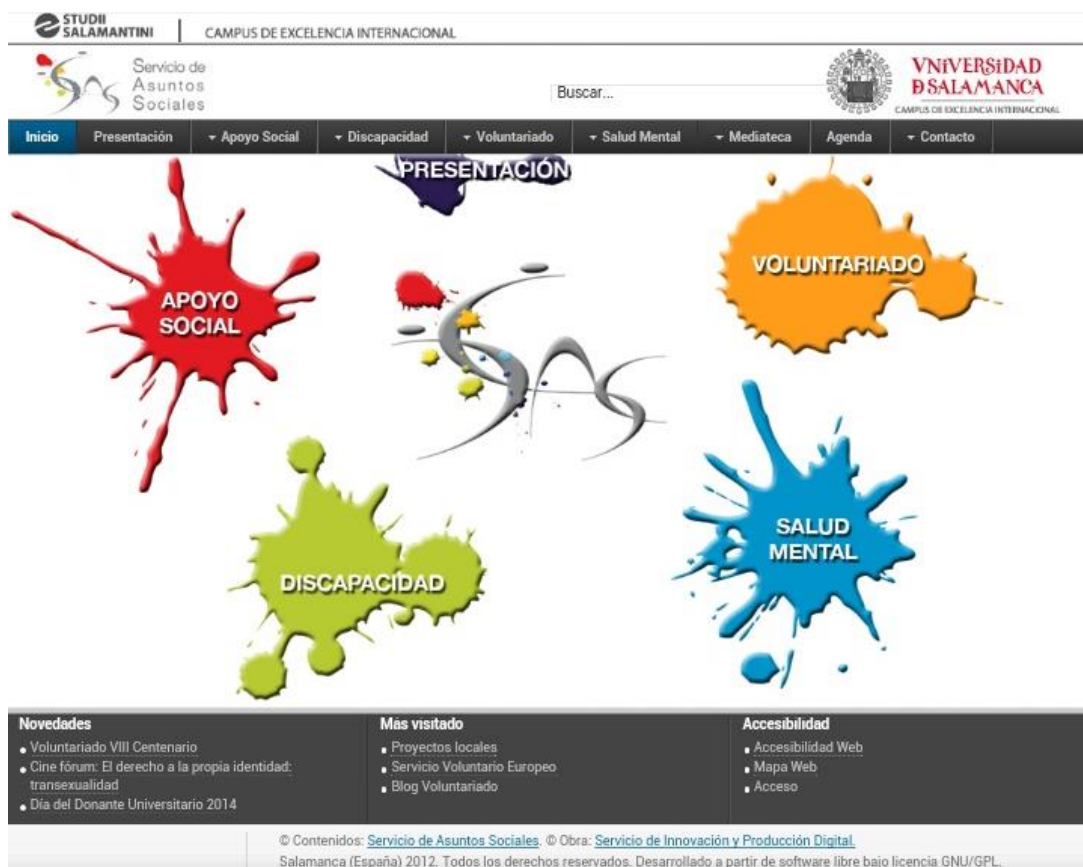


Figura 3. Página de inicio de la antigua página web del Servicio de Asuntos Sociales.



Figura 4. Página de inicio de la actual página web del Servicio de Asuntos Sociales (versión móvil).

Ante el cambio realizado, los responsables de Servicio de Producción e Innovación Digital de la Universidad de Salamanca importaron el contenido de la página web y del blog antiguo al nuevo recurso; así, al margen de la información genérica del SAS, igualmente actualizada, todas las noticias y eventos pasan en Wordpress a integrar la sección de Blog actual, que podemos visualizar en el enlace <http://sas.usal.es/blog/>.

Para organizar los contenidos del Blog, que se remontan a septiembre del 2010, los profesionales del SAS diseñaron una nueva lista de categorías. La adaptación de los contenidos importados a la nueva estructura recayó en el becario de colaboración de la Unidad de Discapacidad. Este proceso se explicará con detalle posteriormente.

Sin embargo, y pese a que las entradas del blog de Diarium habían sido indizadas mediante palabras clave y se contaba con una lista de etiquetas, se optó por no utilizar este recurso ya que se considera que no se había realizado de manera correcta. Esta lista, resultado del trabajo individual de hasta tres becarios del Servicio de Asuntos Sociales en diferentes espacios de tiempo, ha sido considerada por los profesionales como mejorable y ha llevado a que prefieran no utilizar etiquetas en las entradas del blog antes de usar una indización poco profesional.

3. OBJETIVOS

El objetivo general de este Trabajo Fin de Máster viene determinado por la meta marcada por el Servicio de Asuntos Sociales: crear una nueva lista de etiquetas, actualizada y de corte profesional, que facilite la recuperación de información.

De este objetivo general se derivan los siguientes objetivos específicos:

- Realizar un estudio previo para lograr la consecución de una folksonomía adaptada a la actividad de la bitácora del Servicio de Asuntos Sociales.
- Analizar las características del Blog del SAS y más concretamente, su estructura por categorías.
- Evaluar la antigua lista de etiquetas, lo que nos permitirá comprobar la adecuación o no de este recurso a la bitácora del Servicio.
- Analizar las alternativas metodológicas a la hora de indizar los contenidos generados por el Servicio de Asuntos Sociales, teniendo en cuenta las necesidades y características de la información y del propio Servicio.
- Evaluar las diversas alternativas en este caso concreto, comparando las conclusiones con lo dictaminado a nivel teórico por los especialistas.

4. METODOLOGÍA

- **Marco teórico**

Como punto inicial a la hora de realizar este Trabajo Fin de Máster, se ha realizado una revisión bibliográfica, fundamental para escribir el apartado 1 (Marco Teórico). Se ha partido de un estudio general, en el cual se ha consultado documentación sobre la recuperación de información en Internet y el peso de las palabras clave en esta labor, para con posterioridad, centrarse en las diversas modalidades de etiquetado que podemos encontrar en Internet.

En ese sentido, tras la consulta de diversas fuentes bibliográficas, se ha optado por estructurar la exposición de las diferentes opciones fusionando diversas clasificaciones. Se ha partido de la estructura de Hassan Montero (2006) que diferencia indización automática y etiquetado realizado por personas, para en un segundo nivel seguir el modelo planteado por Gómez-Díaz (2012), que diferencia 4 tipos de indización: automática, lenguaje controlado, indización de autores y etiquetado social. Estas dos últimas modalidades, sin embargo, se han integrado en una categoría superior que sería “Lenguaje Natural”, término de uso frecuente en los estudios y que permite realizar una diferenciación clara entre el trabajo realizado a la hora de crear un tesoro y la asignación libre de conceptos, en manos de los creadores como de los lectores de la información.

Dentro de ese mismo apartado, se ha considerado oportuno incluir lo apuntado por Gómez-Díaz, quien se remite a su vez a Javier Cañada (2006) sobre los diversos tipos de indización en función de la metodología y el objetivo planteado; y la diferenciación en función del número de personas, siguiendo de nuevo a Hassan Montero, que diferencia entre personal y colectiva, incluyendo la modalidad de “grupala” ya que en ocasiones el etiquetado está en manos de un grupo limitado de personas, como por ejemplo el caso que nos ocupa con el personal del Servicio de Asuntos Sociales.

Una vez determinada la estructura, se ha optado por realizar un compendio de las ventajas e inconvenientes que cada tipo de modalidad de indización presenta, apuntes que resultan interesantes a la hora de analizar este caso práctico y las conclusiones finales.

- **Estudio de caso**

Como se ha indicado en el marco teórico, las posibilidades a la hora de etiquetar contenidos digitales son muy variadas, con diversos puntos a favor y en contra, por lo que la determinación de una u otra metodología requiere, en primer lugar, una evaluación del recurso; eso implica un análisis del contenido generado y de la organización que se ha establecido o se ha utilizado con anterioridad, tanto a nivel de categorías como de etiquetas. Esto ha permitido conocer de primera mano si el trabajo realizado por los profesionales del Servicio de Asuntos Sociales ha sido el adecuado o si verdaderamente requiere una mejora.

Así, el análisis parte de la descripción de la organización establecida en el antiguo Blog de Diarium a nivel de categorías y la distribución de los artículos. Siguiendo este mismo modelo se ha analizado la estructura actual, para finalmente apuntar a modo de conclusión ciertas mejoras que podrían llevarse a cabo.

A partir de ahí, el trabajo se centró en el estudio del etiquetado del Blog de Diarium, para lo cual resulta esencial realizar un análisis previo de la lista de etiquetas precedentes. Si bien se trata de un recurso descartado por los trabajadores del SAS, su valor para tener un boceto sobre las demandas y usos de información más habituales del centro es innegable. En este sentido, se han descrito sus características, su uso en el antiguo blog, y apuntar ciertos elementos positivos y negativos que esta lista de etiquetas posee.

Posteriormente, se ha realizado el estudio de diversas metodologías de indización que podrían utilizarse en la creación de una futura lista de etiquetas del SAS. Concretamente, se plantearon 3 posibilidades:

- La indización basada en el lenguaje natural, que a su vez podría ser personal, grupal o abiertamente social.
- La indización basada en el lenguaje documental.
- La indización semiautomatizada.

- **Indización basada en el lenguaje natural**

En este caso, se analizó la factibilidad de dicha metodología y los elementos positivos y negativos que cada opción (personal, grupal y social) plantea, en función de lo expuesto por la teoría, la percepción de los profesionales del Servicio de Asuntos Sociales y los problemas planteados en la antigua lista de etiquetas, ya que ellos han utilizado previamente este método. Es importante en este caso tener en cuenta las características de la propia institución ya que si se optase por esta metodología, el diseño y/o la revisión de la lista de etiquetas recaerían en el personal del SAS, al no contar este Servicio con profesionales de la Documentación.

- **Lenguajes documentales**

Se ha seguido el modelo apuntado por Luis Rodríguez Yunta en el artículo del año 2009 *Etiquetado libre frente a Lenguajes Documentales. Aportaciones en el ámbito de Biblioteconomía y Documentación*. En él, el autor recoge las etiquetas utilizadas en hasta tres recursos electrónicos (Documenea, Blogs mantenidos por profesionales de la Documentación y Delicious) y comprueba si estos términos están recogidos en el *Tesaurus de Biblioteconomía y Documentación* (CINDOC, 1995). Este estudio comparativo permite conocer el potencial que los lenguajes documentales tienen en la creación de folksonomías y su adaptación a la realidad digital.

En este caso, el estudio parte de la lista de etiquetas previa del Servicio de Asuntos Sociales y se comprueba su correspondencia en dos lenguajes documentales: el Tesauro de la UNESCO y el Eurovoc. Se ha descartado el uso de tesauros temáticos, ya que las funciones del Servicio de Asuntos Sociales son diversas, por lo que se deberían utilizar varios lenguajes documentales adaptados a diferentes unidades del SAS. Estos tesauros, de corte multidisciplinar, están además disponibles en línea y son por ello actualizables; esto facilita la utilización de este recurso por parte de la población en general y los profesionales del SAS en concreto, además de a priori, garantizar una mayor adecuación a la realidad cibernética.

Así, a partir de la lista de etiquetas, se ha realizado un estudio de los términos recogidos en ambos tesauros, apuntando además a qué microtesauro se adscriben: el hecho de que estos lenguajes documentales sean generales puede llevar a que un mismo concepto pueda ser utilizado en contextos que nada tienen que ver con la actividad del Servicio de Asuntos Sociales. Se analizaron a su vez las claves que determinan por qué el resto de términos no aparecen en dichos tesauros, para en definitiva, descubrir hasta qué punto estos recursos son útiles a la hora de realizar una folksonomía.

- **Indización automática**

Se ha partido de un paquete basado en Java para el procesamiento del lenguaje natural denominado Mallet. Desarrollado por la Universidad de Massachusetts Amherst, es generalmente utilizado para la clasificación de documentos: estableciendo un número de temas, el programa, de forma automática, cataloga los archivos y nos muestra un número x de palabras relevantes que aparecen en los textos de cada tema; ya que nos permite crear un fichero con palabras vacías para que estas sean descartadas, en principio detecta los conceptos que verdaderamente describen a cada grupo.

En este caso concreto, en vez de crear diversos temas o categorías, se ha ordenado al programa la creación de un único tema que contenga numerosos descriptores (en este caso, 299) seleccionados dentro del total de la actividad web del Blog del SAS. De esta manera el programa no cataloga los archivos, sino que los considera a todos como conjunto. Así, teóricamente, se garantiza que los términos más frecuentes en este recurso electrónico estén allí representados, a excepción de los recogidos en el archivo de palabras vacías.

Para ello, se han creado 330 ficheros en formato .txt que recogen cada una de las entradas del blog del SAS, comprendidas entre septiembre del 2010 y julio del 2015³. En cada uno de ellos figura el URL, el título del artículo y el texto de la entrada. Se han eliminado las tildes y demás signos especiales ya que el sistema de Mallet los detecta como error (por ejemplo, Cáritas lo leería como c ritas, descartando la á).

³ Si bien el total de entradas del blog asciende a 385 artículos, muchos de estos son duplicados exactos de otras noticias o tienen enlaces rotos en la actualidad, motivo por el cual han sido descartados.

Para la creación del archivo de palabras vacías, se ha utilizado la lista por defecto contenida en el programa Simple Extractor, un extractor terminológico desarrollado por una spin-off, Dail Software, que tiene su sede en la Universidad Politécnica de Madrid. Se trata de una lista muy completa de términos en español, ya que no sólo incluye preposiciones, pronombres, artículos, adverbios y conjunciones, sino también numerosos verbos de uso frecuente. De nuevo, se ha creado un documento en texto plano con todos estos términos para su detección en Mallet y se han eliminado los caracteres especiales.

El fichero Mallet se ha utilizado gracias al programa de Windows “Símbolo de Sistema”, en el cual se escriben los comandos necesarios. Se importaron los archivos (mallet import-dir), determinando cuál es el archivo de salida de las 330 entradas web (-input) y el archivo de palabras vacías (--stoplist-file). A partir de ahí se fijó el número de temas (--num-topics, en este caso 1) y el número de palabras descriptoras (--num-top-words, en este caso 300, aunque Mallet sólo detectará 299). Igualmente se debieron asignar una serie de archivos de salida de la información, entre los cuales encontramos aquel en el que aparecen los 299 términos utilizados en este análisis (--output-topic-keys).

```

C:\> Símbolo del sistema - mallet train-topics --input import --num-topics 1 --output-topic-keys topic300 --output-doc-topics 300doc --num-top-words 300
Microsoft Windows [Versión 10.0.10240]
(c) 2015 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Rocio>E:
E:\>cd mallet/bin
E:\mallet\bin>mallet import-dir --input archivos --output import --keep-sequence --stoplist-file spanish.txt
Labels =
  archivos
E:\mallet\bin>mallet train-topics --input import --num-topics 1 --output-topic-keys topic300 --output-doc-topics 300doc --num-top-words 300
Data loaded.
Coded LDA: 1 topics, 0 topic bits, 0 topic mask
max tokens: 547
total tokens: 30464
<0> LL/token: -0,31313
<20> LL/token: -0,31313
<30> LL/token: -0,31313
<40> LL/token: -0,31313
0 50 url text title usal http sas voluntariado salamanca h dia personas social actividades universidad programa curso servicio horas informa
al ano mayo asociacion fundacion tu participacion semana derechos discapacidad facultad proyecto jovenes formacion noviembre educacion campus marzo cent
royectos solidario jornadas espana colaboracion dias asuntos viernes castilla accion c puedes intergeneracional jornada zamora anos concurso oxfam inter
estudiantes ninos teatro com grupo apoyo web objetivo voluntaria forma correo juventud encuentro octubre comunidad taller solidaria cine san sabado orga
alud vida i talleres entrada lunes junio ciudad tormes actos entidades electronico edicion celebra mayores universitario charla pais miercoles concierto
a voluntarias sanchez donacion julio horario poblacion sangre v celebrara oficina experiencia asociaciones diciembre derecho mesa unidas realizara organ
o pagina premio mundial local aula quieres pretende general contacto sensibilizacion espacio organizacion cabo punto mes participantes recursos superior
madrid situacion organizaciones domingo presentacion formulario red dna favor motivo cafe igualdad iniciativa nacional espanola lucha cambio ciclo for
ciudadania crisis objetivos jose fonseca intercultural as s medula vih pyfano fecha mujeres dirigido pasos carrera marcha provincia datos celebracion
mercio conferencia aviva colegio nivel gratuita sve emprendimiento animacion prevencion mujer santa genero reflexion paz asistencia beneficio acto medic
donantes innovation radio futuro consumo numero plazas promocion entradas especial nombre informativa junta servicios iii grupos europea propuestas juv
<50> LL/token: -0,31313
<60> LL/token: -0,31313
<70> LL/token: -0,31313
<80> LL/token: -0,31313
<90> LL/token: -0,31313

```

Figura 5. Funcionamiento del clasificador documental Mallet en el programa Símbolo de Sistema.

Se obtuvo una lista de términos frecuentemente utilizados por los profesionales del SAS en las entradas del blog, pero el uso habitual de un concepto no implica necesariamente que sea un buen descriptor, por lo que en este estudio se analizaron las características de estos términos. Se han evaluado, en definitiva, los pros y contras que el uso de una herramienta de estas características puede tener a la hora de designar descriptores web y para la creación de una folksonomía.

A modo conclusivo, se evaluarán las ventajas y limitaciones de los sistemas de indexación aplicados a este caso práctico y cuál podría ser la solución más recomendable a la hora de crear una lista de etiquetas para el Blog de esta institución.

5. ANALISIS DE CASO

5.1. La categorización del Blog del Servicio de Asuntos Sociales

- **La categorización en el Blog de Diarium**

El Blog de Diarium del SAS, al que a día de hoy no podemos acceder pero cuya URL era <http://diarium.usal.es/voluntariadosas>, contaba con 284 entradas web, repartidas en 7 categorías diferentes.

- Colaboradoras
- Eventos
- Formación
- General
- Participación
- Sensibilización
- SVE / EVS / Voluntariado Europeo

Resulta llamativo que tanto el nombre de la URL del blog como algunas de las categorías hacen referencia concreta a la actividad de la Unidad de Participación y Voluntariado. Por tanto, no se puede considerar a este blog como verdaderamente representativo del conjunto del SAS, aunque en categorías genéricas como *Eventos* o *Sensibilización* pudiese informarse de acciones de otras secciones de la institución.

A modo informativo, cabe señalar que *General* era una categoría indefinida, en la que se incluían noticias de todo tipo; *Colaboradoras* hace referencia a las entidades externas con las que se mantienen acuerdos de colaboración y *Sensibilización* resaltaría las campañas organizadas por las diferentes unidades.

Esta categorización mezcla lo que serían acciones (el Servicio Voluntario Europeo) con tipos de actos (eventos, formación), prescindiendo además de una organización que diferencie a cada sección del SAS. Se trata, por tanto, de un modelo caótico y poco representativo de la institución.

En cuanto a la distribución de los artículos por categorías, se comprueba como muchos de ellos han sido ubicados en varias al mismo tiempo, ya que la suma total (388) es mayor al número de artículos (284). Las noticias generadas por la Unidad de Participación y Voluntariado son, según estas cifras, las principales del recurso, con 107 noticias vinculadas a *Participación*, que se unen a las 8 de *SVE / EVS / Voluntariado Europeo* y a las 37 de *Colaboradoras*, ya que es esta sección la que establece relaciones con otras asociaciones de ámbito local y regional. No se puede descartar su presencia en otras categorías de corte genérico, por lo que en definitiva, la imagen que transmite el recurso es de estar focalizado en la promoción de las acciones desarrolladas por una unidad y no del conjunto del SAS.

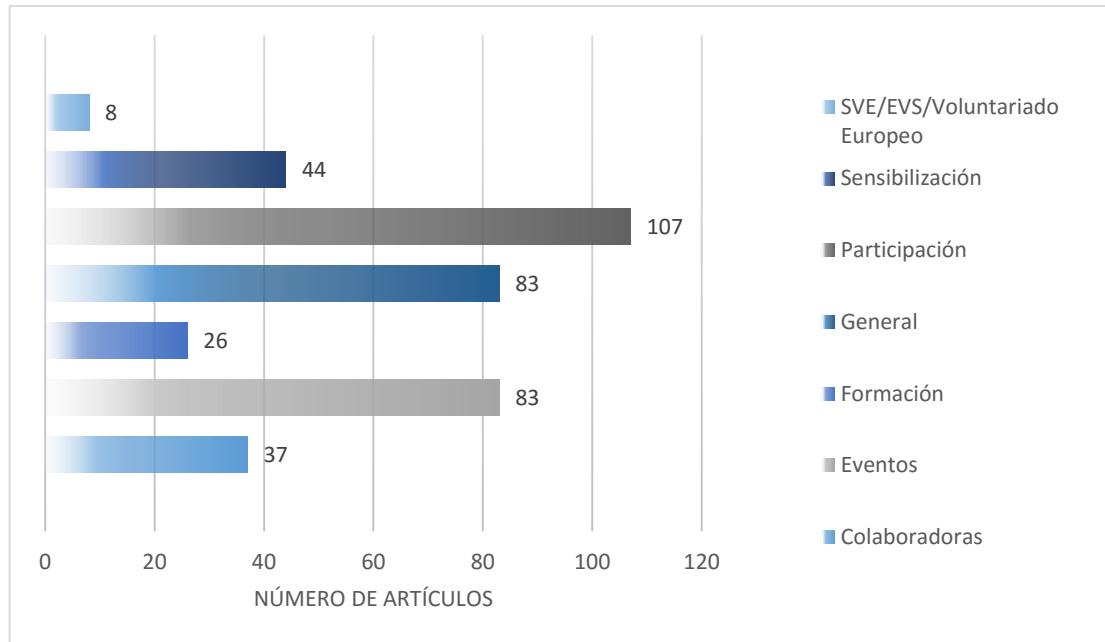


Figura 6. Distribución de las publicaciones por categorías en el Blog de Diarium del SAS.

5.2. La categorización en el Blog actual

Con la creación de la nueva página, se realizó una nueva categorización de los contenidos del Blog, más compleja y que ahora sí, representa a todos los servicios y tipos de actos del Servicio de Asuntos Sociales.

De entrada, se divide en dos niveles principales, el de Áreas y el de Secciones. Áreas incluye las categorías en función de las diversas unidades del Servicio, mientras que Secciones permite la identificación de la noticia en función del tipo de actividad o la naturaleza de la publicación.

Así, en la sección de Áreas nos encontramos la siguiente subdivisión:

- 1) Apoyo Social
 - i) Programa Intergeneracional
- 2) Cooperación al Desarrollo
- 3) Discapacidad
- 4) Participación y Voluntariado
 - i) Servicio Voluntario Europeo
- 5) Psicología y Psiquiatría
- 6) Usaludable
 - i) Drogas
 - ii) Otras adicciones
 - iii) Sexualidad
 - iv) Violencia interpersonal

Se puede observar que no hay una equivalencia exacta entre la estructura del Servicio de Asuntos Sociales y esta categorización. Las cuatro primeras corresponden a las unidades y las subcategorías a programas semiindependientes (en el caso del Servicio Voluntario Europeo no es así, pero se justifica por el interés específico en este programa y porque se gestiona de manera diferente al voluntariado a nivel local).

Sin embargo, la actividad de dos unidades como son la de Psiquiatría y la de Psicología se integran en una sola mientras que un programa dependiente (USALudable) tiene una categoría propia con hasta 4 subcategoría. Al mismo tiempo, el proyecto desarrollado en Ávila con los jóvenes con dificultades del lenguaje carece de sección o subsección. Los responsables de la página del Servicio de Asuntos Sociales justifican esta estructura basándose en el hecho de que estas unidades prestan sus servicios en las facultades (sea de Psicología, Medicina o en la Escuela de Educación y Turismo de Ávila) y no suelen realizar actividades fuera de ese contexto, por lo que tampoco se suelen generar noticias entorno a estas unidades. Sin embargo, el responsable del *Programa Usaludable* realiza campañas de sensibilización en las diversas temáticas anteriormente citadas, lo que lleva a que sea esencial informar sobre estas actividades en el blog del Servicio de Asuntos Sociales.

En cuanto a las Secciones, los responsables del Servicio han diferenciado los recursos en diversas categorías:

- Campañas
- Consultas
- Destacados
- Eventos
- Formación
- Mediateca
- Noticias
- Participa

Como se puede observar, se trata de un cajón de sastre en el cual se integran categorías descriptivas de actividades, como son campañas, eventos o formación con otras: la sección de *Consultas* nace con la finalidad de dar respuesta a dudas que los usuarios puedan plantear; *Destacados* permite la inclusión de las entradas más relevantes dentro de la página principal del Servicio; se incluiría en *Mediateca* todo el contenido visual o sonoro como es el caso de vídeos, audios (por ejemplo, del programa de Radio USAL “Dejando Huella”) y las menciones realizadas en medios sobre las acciones del Servicio de Asuntos Sociales; por *Noticias* toda aquella novedad relevante tanto del propio organismo como de otros (colaboradores o no) y la función de *Participa* sería la de dar visibilidad a campañas y eventos en los que la comunidad universitaria puede colaborar de forma activa.

La mayoría de entradas tienen categorías múltiples; por ejemplo, un evento en el cual participan varias unidades es susceptible de estar vinculado a cada una de ellas, al mismo tiempo que lo está a la sección *Eventos* y en *Destacados* si interesa que figure en la página principal. Igualmente, las noticias asociadas a subcategorías pueden estar vinculadas a la categoría raíz, aunque no necesariamente (por ejemplo, el Programa Intergeneracional tiene entidad propia y no tiene porqué asociarse a Apoyo Social).

Si analizamos los artículos asociados a cada categoría en la actualidad (sumando tanto las entradas de la página web antigua como los realizados ya en la nueva interfaz) el número total debería ser superior al número de artículos (la muestra total es de 385 entradas web⁴).

La distribución del apartado de Áreas por número de artículos sería la siguiente:

CATEGORÍAS	NÚMERO ARTICULOS
APOYO SOCIAL	30
PROGRAMA INTERGENERACIONAL	15
COOPERACIÓN AL DESARROLLO	42
DISCAPACIDAD	22
PARTICIPACIÓN Y VOLUNTARIADO	85
SERVICIO VOLUNTARIO EUROPEO	19
PSICOLOGÍA Y PSIQUIATRÍA	2
USALUDABLE	24
DROGAS	12
OTRAS ADICCIONES	1
SEXUALIDAD	5
VIOLENCIA INTERPERSONAL	1
TOTAL	258

Tabla 1. Distribución de las publicaciones por categorías de Áreas en el Blog actual del SAS.

⁴ Fecha de consulta: 24 de julio de 2015.

Como se puede observar, el número total de artículos vinculados a un área es notablemente inferior al total de entradas. Esto se debe a que la mayoría de artículos importados de Diarium han sido únicamente categorizados por Secciones y no por Áreas. Las entradas creadas en la nueva interfaz han sido asociadas a los diferentes servicios del SAS, pero en cuanto al resto no hay una política definida, por lo que hay contenidos del año 2010 bien categorizados al lado de otros que no lo están. Por ello, los datos deben ser considerados como meramente orientativos a la hora de analizar el peso que tienen en el blog las diferentes unidades del Servicio de Asuntos Sociales.

Igualmente, se puede observar que la mayoría de servicios cuentan con una actividad similar, de entre 20 y 40 artículos, a excepción de la descompensación entre *Participación* y *Voluntariado* (85) con respecto a *Psicología* y *Psiquiatría*, con 2, que demuestra la escasa actividad de ciertas unidades con respecto a otras. Otro elemento llamativo es el hecho de que la suma de las subcategorías de *USALudable* (19) es inferior al de la categoría madre (24), por lo que se podría deducir que hay actividades que están doblemente categorizadas y otras que no lo están.

En cuanto a las Secciones, los resultados por categoría pueden ser visualizados en el siguiente gráfico:

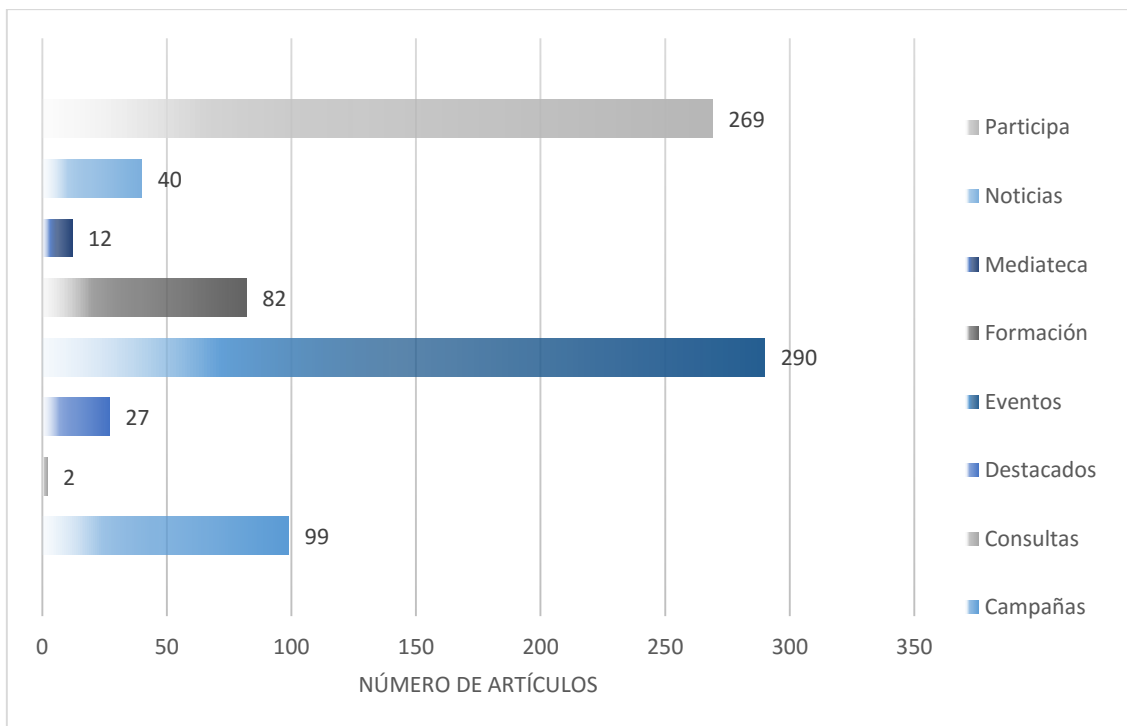


Figura 7. Distribución de las publicaciones por categorías de Secciones en el Blog actual del SAS.

Se observa que los artículos del Blog han sido asignados al menos a una de estas categorías, ya que la suma total en este caso asciende a las 821 menciones, por lo que se puede incluso afirmar que las entradas cuentan en muchos casos con más de una categoría de este apartado. Igualmente, se puede observar una desigualdad aún más notable que en el caso de las Áreas: aquí, mientras Participa o Eventos suman casi 300 artículos (269 y 290 respectivamente), Consultas no pasa de las dos entradas, o Mediateca de 12.

A la hora de analizar la categorización y la adecuación al contenido real de las noticias, se han observado errores notables en las Secciones. Por ejemplo, los dos artículos de *Consultas* son en realidad eventos, y por tanto no nacen de las dudas de los usuarios; igualmente, se ha considerado que en *Mediateca* deben aparecer los eventos de Cine Fórum (ciclos de películas sobre temáticas tratadas en Asuntos Sociales), cuando en teoría se reserva a la publicación de contenido visual y sonoro.

Estos problemas a la hora de la categorización de los contenidos se pueden asociar a que, como se mencionó anteriormente, la asignación de las antiguas entradas a la nueva estructura de categorías recayó en exclusiva en el becario de colaboración de la Unidad de Discapacidad, quien carecía de una estructura semántica sobre la que basar la categorización, más allá de la representación de la jerarquía orgánica del Servicio.

5.3. El etiquetado del Blog del Servicio de Asuntos Sociales

- **La antigua lista de etiquetas del Blog de Diarium**

El etiquetado en un blog como el del Servicio de Asuntos Sociales resulta muy relevante a la hora de facilitar la recuperación de información. Como se ha visto en la descripción del organismo, cada una de las unidades del SAS realiza un número notable de actividades diversas, por lo que poder localizar las entradas en función de una serie de términos más allá de la categorización por Áreas o Secciones facilitaría notablemente la navegación tanto de los usuarios como de los propios profesionales.

Como se indicó anteriormente, en el antiguo Blog de Diarium se contaba con una lista de etiquetas. Cuando se realizó la importación de los artículos a la nueva página web (un total de 284 entradas, lo que supone el 73,7% del total actual) se importó igualmente la lista de etiquetas, compuesta por 193 términos; sin embargo, los profesionales del SAS optaron por descartar este recurso en su nueva página de Wordpress. Las claves de esta decisión han sido las siguientes:

- El escaso volumen de etiquetado. La mayor parte de las palabras clave han sido utilizadas una única vez (135, un 70% del total), habiendo incluso descriptores que nunca han sido utilizados (12). La etiqueta más frecuente, *ambiental*, sólo se menciona en 17 ocasiones, lo que resulta escaso teniendo en cuenta el volumen de entradas. Así, la distribución porcentual de las etiquetas en función del número de artículos a las que se asignan sería la siguiente:

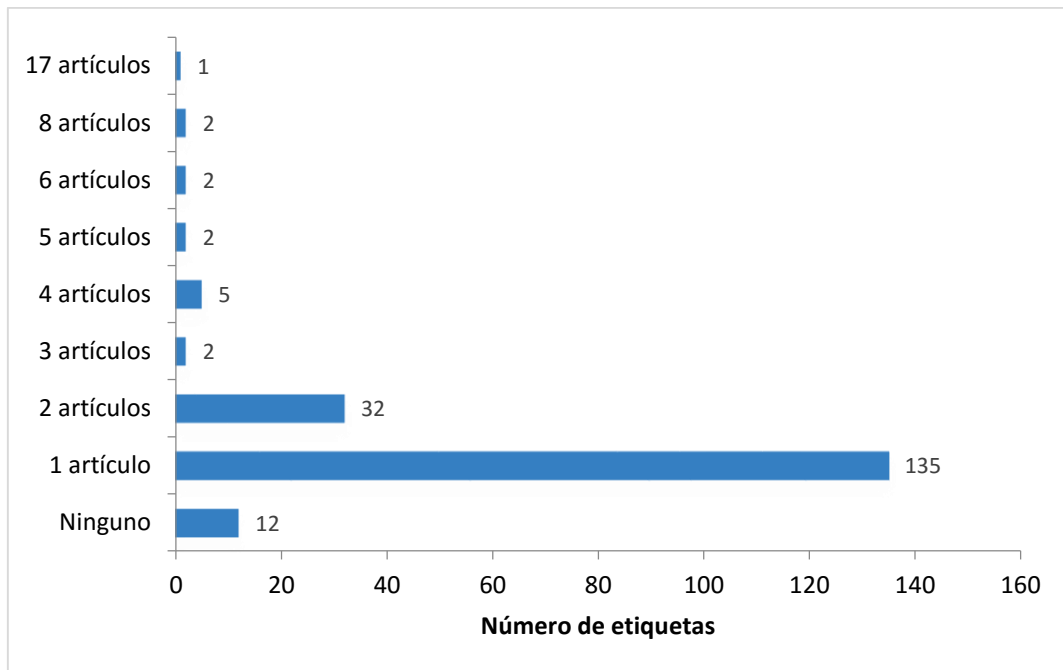


Figura 8. Distribución de las etiquetas por número de artículos en el Blog Diarium del SAS.

Algunas de las claves que explicarían el escaso uso de las etiquetas se pueden encontrar en los siguientes puntos.

- Faltas de ortografía y errores tipográficos. Algunos descriptores se encuentran de forma duplicada, en la cual una de las formas es incorrecta (por ejemplo, se cuenta con la etiqueta *cooperación* y con una denominada *cooperaci*); términos que nada tienen que ver con la actividad del SAS (*malakai*) o incluso símbolos, como]. alguna de estas etiquetas nunca han sido utilizadas, pero otras sí lo han hecho. Si eliminásemos de la lista los que son considerados como errores, se reduciría a 187 términos.
- Duplicidades. Suele ser la realidad más habitual, lo que explicaría que muchos términos sean utilizados una única vez. No se pueden considerar como erratas, pero son claras redundancias que provocan una dispersión de la información. Algunos ejemplos serían:
 - *Coordinadora de ONGD / Coordinadora ONG para el desarrollo / Condg*
 - *Fundación Tormes / Fundación Tormes – EB*
 - *Intermón Oxfam / Oxfam Intermón*
 - *Red de voluntariado social / Red Salamanca*
 - *Café Tertulia / Tertulia Café*

Otras veces las duplicidades no son tan evidentes, pero son términos que para los responsables del Servicio de Asuntos Sociales vendrían a describir lo mismo. Contamos con ejemplos como:

- *Campus inclusivos / Campus sin límites*
 - *Ciclo de cine / CineForum* (la denominación interna del SAS)
 - *Cooperación / Cooperación al desarrollo*
- Hay ejemplos de varias etiquetas que comparten una misma palabra raíz, tras la cual figuran diversos adjetivos específicos. En algunos casos son términos que se podrían describir de forma independiente.

Por ejemplo, *Derechos* y *Derechos humanos* son dos etiquetas que no necesariamente son incompatibles; *Derechos* como concepto puede estar asociado a la palabra *Niños*, *Mujer*, o *Indígenas* (de hecho, existe la etiqueta *Derecho indígenas*).

Sin embargo, ante dos etiquetas como son la de *Niños* y la de *Niños con cáncer*, la segunda podría eliminarse y optar por el uso de dos etiquetas como son *Niños* y *Cáncer* (una palabra que, además, no aparece en este listado). Ejemplos similares serían:

- *Concurso / Concurso de cortos*
 - *Emprendimiento Cultural / Emprendimiento Social*
 - *Viaje / Viajes solidarios*
- Palabras derivadas. Se han encontrado ejemplos en los que las etiquetas figuran tanto en singular como en plural (*curso / cursos*). En otros casos, pertenecen a la misma familia, por lo que se podría optar por un mismo concepto en todos los casos, tal es el caso de:
 - *Empleabilidad / Empleo*
 - *Inmigración / Inmigrantes*
 - *Tejer / Tejidos*
 - Conceptos recogidos en la antigua categorización, por lo que su utilización como etiqueta resultaría redundante. Tal es el caso de *Participación*, *Sensibilización*, *SVE/EVS/Voluntariado Europeo* y *Formación*.

Como se puede observar, la mayor parte de los inconvenientes habituales del etiquetado en lenguaje natural están presentes en menor o mayor medida en esta lista de etiquetas. El hecho de que no fuese realizada de forma consensuada sino entre distintos agentes de manera intermitente contribuye a que el número de duplicidades sea abultado, ya que cada uno de ellos ha podido denominar un mismo concepto de manera diferente sin que la lista final fuese revisada. La falta de interés por parte de los profesionales del Servicio en cuanto a la supervisión de los conceptos utilizados y a la utilización de estos para el etiquetado de los artículos evidencia el carácter puntual de esta labor.

Sin embargo, se encuentran elementos positivos en esta folksonomía. El etiquetado basado en el lenguaje natural tiene como principal ventaja la adaptación de la etiqueta a la realidad descrita, lo que se puede comprobar gracias al nivel de especificación temática y local de las etiquetas utilizadas en este caso concreto.

Así, a partir de la lista de 187 términos (aquella en la que se han eliminado los conceptos considerados claramente como erratas, pero se mantienen las duplicidades), podemos comprobar cómo un número notable de palabras clave hacen referencia a eventos con denominaciones muy específicas organizados por el Servicio de Asuntos Sociales (etiquetas como “formación básica” o “curso de cortos” no se incluyen ya que pueden ser utilizadas en amplios contextos), al igual que a nombres de asociaciones, personas, ayuntamientos de la zona y otras entidades locales.

TIPOS DE NOMBRES PROPIOS	NÚMERO ETIQUETAS
ASOCIACIONES LOCALES	25
ASOCIACIONES NACIONALES	13
AYUNTAMIENTOS DE LA ZONA	3
EVENTOS OTRAS ASOCIACIONES (TÉRMINO PROPIO)	4
EVENTOS SAS (TÉRMINO PROPIO)	6
OTRAS ENTIDADES LOCALES	2
PERSONAS	1
TOTAL	54

Tabla 2. Número de etiquetas asociadas a nombres propios en la folksonomía creada por el SAS.

Por tanto, casi un 30% de los términos hacen referencia a nombres propios asociados a realidades muy específicas del Servicio de Asuntos Sociales y su ámbito de acción. Por ello, podemos afirmar cómo una parte notable del etiquetado del SAS es de corte “amiguista”, ya que ciertos términos no son de conocimiento público, pero sí fácilmente identificables tanto para los profesionales como para muchos de los usuarios del Blog.

De igual manera, no hay ejemplos de etiquetas “egoístas” como cabría pensar en un indizado personal, por lo que se ha tenido en cuenta siempre los intereses de la institución. Mejorar el etiquetado, utilizando en la medida de lo posible etiquetas “altruistas” y evitando las populares, poco descriptivas y profesionales, sería recomendable de cara a un futuro.

5.4. Propuesta y valoración de las modalidades de etiquetado web

5.4.1. Etiquetado realizado por personas en lenguaje natural

La realización de una folksonomía a partir de los términos elegidos por los autores y usuarios de la información digital sería de entrada la opción más adaptada a la realidad de Internet. De todos modos, se deben de tener en cuenta notables inconvenientes, como los evidenciados en la antigua lista de etiquetas. Contamos con tres posibilidades: el etiquetado personal, el grupal y el social.

Siguiendo la estructura planteada por los especialistas, a la hora de realizar este análisis se ha tenido en cuenta la realidad y la percepción de los profesionales del Servicio de Asuntos Sociales y los problemas planteados en la antigua lista de etiquetas, para así determinar cuál de estas modalidades se adaptaría mejor a las características del recurso.

- **Etiquetado personal**

Como se ha comprobado en el caso de la anterior folksonomía, la opción de la indización por parte de una única persona no resulta la opción más adecuada. Si bien se simplificaría notablemente la labor del etiquetado, al tratarse de una web profesional en la que se ofrece un servicio a la comunidad, la información e indización de los contenidos debe ser clara para los usuarios. Cierto es que la antigua lista no tiene un carácter egoísta, pero la falta de supervisión por parte de otras personas puede llevar a la larga a la utilización de términos poco adecuados, redundancias dentro de la lista, entre otros problemas ya expuestos.

En el caso de que se hiciese un etiquetado personal, se recomendaría que fuese una labor realizada por alguno de los profesionales del Servicio, debido al grado de conocimientos que este puede tener sobre el conjunto de la institución. Por ello, dejar esta responsabilidad en manos de becarios como se ha realizado hasta ahora no sería lo más recomendable.

Aun así, se podrían evidenciar desigualdades entre unidades a la hora de la descripción, ya que el profesional conocerá mejor su rango de acción que el de otros servicios del SAS, lo que permitiría ser más específico en unas noticias que en otras. Al mismo tiempo, el encargado de esta labor podría utilizar términos que no son de uso frecuente para el responsable de cada unidad, que se ocupa de redactar los artículos que atañen a su campo de acción.

- **Etiquetado grupal**

El trabajo colectivo de indización en manos de todos los responsables del Servicio de Asuntos Sociales podría ser una buena opción, teniendo en cuenta que el personal de cada unidad es el que mejor conoce la realidad a la que hace frente.

Al mismo tiempo, son ellos los que redactan las entradas; por tanto, sería recomendable que fuesen los que realizasen en un primer nivel una lista de etiquetas de su ámbito de acción, para posteriormente realizar un trabajo común en el cual se podrían discutir los términos a utilizar. Igualmente, se debería organizar la revisión periódica de las etiquetas para evitar redundancias, conceptos obsoletos, etc. Esto obliga, al igual que pasaría en el caso de la categorización, al establecimiento de una política orientada al etiquetado del Blog.

Sin embargo, a diferencia de lo que ocurre con las categorías (donde al menos se cuenta con una estructura definida), los profesionales del SAS nunca han realizado esta labor y ello podría llevar a que a la larga se pudiesen evidenciar problemas similares a los del etiquetado personal. A esto se añade que la realización desde el principio de una lista de etiquetas requiere de un trabajo de revisión exhaustivo difícilmente asumible por el personal del Servicio, fundamentalmente por una cuestión de tiempo.

- **Etiquetado social**

La creación de una folksonomía que nazca de los descriptores que todos los usuarios del blog del Servicio de Asuntos Sociales utilizarían para identificar los artículos se considera poco viable y adecuada. Si bien se garantizaría que los lectores recuperasen la información con mayor facilidad, ya que serían ellos los que determinan cómo se etiqueta, aplicar este modelo en una página web o blog resulta complejo. Los blogs no están diseñados para dar la opción al usuario de indizar el contenido, ya que es una responsabilidad del autor, a diferencia de lo que ocurre en redes sociales u otros recursos electrónicos, dotados muchos de ellos de apartados en los que se invita al usuario a determinar la/s palabra/s más adecuada/s para describir el contenido. Esta limitación tecnológica determina que sea imposible hablar de etiquetado social en este caso; si bien se podría involucrar a personas ajenas al SAS para determinar las etiquetas a utilizar, no es una funcionalidad abierta a todo potencial lector.

Además, se trata de una página de corte institucional; esto implica que la selección de las etiquetas debe ser, en última instancia, una responsabilidad de los profesionales, que son los que conocen los términos apropiados; si bien los usuarios tendrían la capacidad de revisar el uso de las etiquetas, podrían encontrarse casos de “etiquetado populista”, con conceptos poco descriptivos y adecuados.

Aplicar un método mediante el cual un grupo determinado de personas, ajenas al Servicio de Asuntos Sociales, determinen los términos más adecuados para todo el conjunto de artículos del Blog resulta complicado de llevar a la práctica por los motivos señalados en el caso del etiquetado grupal: la falta de experiencia y de tiempo de los profesionales para coordinar una tarea de estas características, unido al volumen importante de artículos a revisar, que dificulta el involucramiento de voluntarios en esta tarea.

5.4.2. Lenguajes documentales

Si bien la creación de un tesoro propio para la diversa actividad del Servicio de Asuntos Sociales resulta inasumible en el momento actual, la consulta de los lenguajes documentales a la hora de indizar contenidos asegura la adecuación de los conceptos, y por tanto, se alcanzaría uno de los objetivos perseguidos por el SAS: el uso de una terminología correcta y profesional a la hora de etiquetar en el Blog.

Sin embargo, la diversidad de funciones que abarca el SAS lleva a que sea complicado encontrar un único tesoro que pueda abarcar todo el Servicio, desde campos tan diferentes como son la Psiquiatría o la Cooperación Internacional. De todos modos, se pueden encontrar tesoros temáticos en español centrados en la realidad de cada unidad o la de varias. Por citar algunos ejemplos, tendríamos:

- *Tesoro de Servicios Sociales* (IMSERSO, 1986);
- *Tesoro de Trabajo Social* (Dávila, 2005);
- *Tesoro en Ciencias Sociales* (Matamoros, 2001);
- *Tesoro ISOC* (Instituto de Estudios Documentales sobre Ciencias y Tecnología, 1995), centrado en diversas temáticas; en lo que respecta al Servicio de Asuntos Sociales, hay tesoros Psiquiatría, Sociología o Drogas.
- *Tesoro de la Discapacidad* (Observatorio Estatal de la Discapacidad, 2009), esencial en los asuntos relativos a este particular y que puede abarcar incluso asuntos de otras Unidades del SAS.

Para este estudio, se ha optado por el uso de tesoros genéricos, actualizados y disponibles en línea. Dos de los más conocidos son el Tesoro de la UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) y el Eurovoc. La lista controlada de la UNESCO recoge hasta 9000 términos en español sobre educación, cultura, ciencias naturales, sociales y humanas, además de la comunicación y la información. Mientras, el Eurovoc es el tesoro multilingüe de la Unión Europea; impulsado por iniciativa del Parlamento Europeo entre otros, nace con el fin de abarcar todas las áreas de actividad de las instituciones comunitarias.

Siguiendo la metodología de Rodríguez Yunta (2009), en la cual compara una lista de etiquetas con los términos de un tesoro, se ha considerado oportuno para analizar la potencialidad de las listas controladas hacer una comparación entre el etiquetado realizado hasta ahora por el Servicio de Asuntos Sociales y su correspondencia en estos tesoros generales, analizando simplemente que términos aparecen independiente de su rango (término genérico, específico, relacionado, etc.).

Si partimos de la lista revisada de 187 términos, cabe señalar de nuevo que 54 son nombres propios o hacen referencia a realidades locales y concretas del SAS. Generalmente estos términos no aparecen nunca en los tesoros, y en este caso en concreto, hay una sola excepción: "Amnistía Internacional" aparece como término en el Tesoro Eurovoc como ejemplo de organización internacional no gubernamental. Así, podemos decir que al menos 53 de los 187 términos (28,34%) no son recogidos en el tesoro por tratarse de identificadores.

En el Tesouro Unesco aparecen 35 términos que se corresponden exactamente con las etiquetas del antiguo Blog del SAS, mientras que en el Eurovoc podemos hablar sólo de 29. Veintiuna palabras clave aparecen en ambos lenguajes documentales. De todos modos, la equivalencia exacta representa solamente un 18,71% (Unesco) y un 15,5% (Eurovoc) de la totalidad de etiquetas existentes. La lista sería la siguiente:

UNESCO	EUROVOC
AMAZONIA	
	AMNISTÍA INTERNACIONAL
AYUDA HUMANITARIA	AYUDA HUMANITARIA
	AYUNTAMIENTO
CAMBIO SOCIAL	CAMBIO SOCIAL
CAMBOYA	CAMBOYA
	CASTILLA Y LEÓN
CATÁSTROFE	
COMUNICACIÓN	COMUNICACIÓN
CONFERENCIA	
	COOPERACIÓN AL DESARROLLO
CORAZÓN	
CURSO	
DERECHOS HUMANOS	DERECHOS HUMANOS
DESIERTO	DESIERTO
DISCRIMINACIÓN	DISCRIMINACIÓN
	DONACIÓN
EMPLEO	
FILIPINAS	FILIPINAS
FORMACIÓN	

UNESCO	EUROVOC
FORMACIÓN BÁSICA	
INFANCIA	INFANCIA
INMIGRACIÓN	INMIGRACIÓN
JUSTICIA SOCIAL	
JUVENTUD	JUVENTUD
	LANA
MARRUECOS	MARRUECOS
MEDIO AMBIENTE	
NIGERIA	NIGERIA
NOTICIAS	
PARAGUAY	PARAGUAY
PARTICIPACIÓN	PARTICIPACIÓN
POBREZA	POBREZA
RACISMO	RACISMO
RADIO	RADIO
SALUD	
SECUESTRO	SECUESTRO
SIRIA	SIRIA
	TURISMO RESPONSABLE
UNIVERSIDAD	UNIVERSIDAD
VIAJE	VIAJE
VOLUNTARIADO	

Tabla 3. Etiquetas de la folksonomía creada por el SAS presentes en los tesauros UNESCO y Eurovoc.

Como se puede observar en esta lista, las equivalencias son de todo tipo: por cuestiones geográficas (zonas, países o regiones); por aspectos relacionados con la comunicación (“radio”, “noticias”), o conceptos sin directa relación con las temáticas del SAS como es el caso de “Iana”; aun con todo, hay una presencia importante de términos verdaderamente significativos con respecto a los campos de acción del SAS. Para analizar mejor este aspecto, se ha realizado una comparativa para conocer los microtesauros a los que están adscritos estos términos en cada uno de los dos lenguajes documentales.

ETIQUETA	MICROTESAURO UNESCO	MICROTESAURO EUROVOC
AMAZONIA	AGRUPACIONES GEOGRÁFICAS	
AMNISTÍA INTERNACIONAL		ORGANIZACIONES NO GUBERNAMENTALES
AYUDA HUMANITARIA	RELACIONES INTERNACIONALES	POLÍTICA DE COOPERACIÓN
AYUNTAMIENTO		PODER EJECUTIVO Y ADMINISTRACIÓN PÚBLICA
CAMBIO SOCIAL	SISTEMAS SOCIALES	VIDA SOCIAL
CAMBOYA	ASIA Y PACÍFICO	ASIA – OCEANÍA
CASTILLA Y LEÓN		REGIONES DE LOS ESTADOS MIEMBROS DE LA UE
CATÁSTROFE	POLUCIÓN, CATÁSTROFES Y SEGURIDAD	
COMUNICACIÓN	INVESTIGACIÓN Y POLÍTICA DE LA COMUNICACIÓN	COMUNICACIÓN
CONFERENCIA	INVESTIGACIÓN Y POLÍTICA DE LA COMUNICACIÓN	
COOPERACIÓN AL DESARROLLO		POLÍTICA DE COOPERACIÓN
CORAZÓN	BIOLOGÍA	
CURSO	GESTIÓN DE LA EDUCACIÓN	

ETIQUETA	MICROTESAURO UNESCO	MICROTESAURO EUROVOC
DERECHOS HUMANOS	DERECHOS HUMANOS	DERECHOS Y LIBERTADES
DESIERTO	GEOGRAFÍA Y OCEANOGRAFÍA	MEDIO NATURAL
DISCRIMINACIÓN	PROBLEMAS SOCIALES	DERECHOS Y LIBERTADES
DONACIÓN		DERECHO CIVIL
EMPLEO	TRABAJO	
FILIPINAS	ASIA Y PACÍFICO	ASIA – OCEANÍA
FORMACIÓN	ENSEÑANZA Y FORMACIÓN	
FORMACIÓN BÁSICA	ENSEÑANZA Y FORMACIÓN	
INFANCIA	POBLACIÓN	DEMOGRAFÍA Y POBLACIÓN
INMIGRACIÓN		MOVIMIENTOS MIGRATORIOS
JUSTICIA SOCIAL	POLÍTICA Y BIENESTAR SOCIAL	
JUVENTUD	POBLACIÓN	DEMOGRAFÍA Y POBLACIÓN
LANA		PRODUCTOS DE ORIGEN ANIMAL
MARRUECOS	ÁFRICA	ÁFRICA
MEDIO AMBIENTE	CIENCIAS AMBIENTALES E INGENIERÍA	
NIGERIA	ÁFRICA	ÁFRICA

ETIQUETA	MICROTESAURO UNESCO	MICROTESAURO EUROVOC
NOTICIAS	INDUSTRIA DE LA INFORMACIÓN	
PARAGUAY	AMÉRICAS Y CARIBE	AMÉRICA
PARTICIPACIÓN	SISTEMAS SOCIALES	ORGANIZACIÓN DE LA EMPRESA
POBREZA	PROBLEMAS SOCIALES	CONTABILIDAD NACIONAL
RACISMO	CUESTIONES ÉTNICAS	DERECHOS Y LIBERTADES
RADIO	INDUSTRIA DE LA INFORMACIÓN	INDUSTRIAS NUCLEAR Y ELÉCTRICA
SALUD	CIENCIAS MÉDICAS	
SECUESTRO	PROBLEMAS SOCIALES	DERECHO PENAL
SIRIA	ASIA Y PACÍFICO	ASIA – OCEANÍA
TURISMO RESPONSABLE		VIDA SOCIAL
UNIVERSIDAD	ESTABLECIMIENTOS DE ENSEÑANZA	ENSEÑANZA
VIAJE	SERVICIOS	VIDA SOCIAL
VOLUNTARIADO	POLÍTICA Y BIENESTAR SOCIAL	

Tabla 4. Distribución de las etiquetas de la folksonomía del SAS en microtesauros UNESCO y Eurovoc.

Así, se demuestra cómo si bien la mayoría de los términos y microtesauros tienen una directa relación con la actividad del SAS, e incluso es habitual la correspondencia entre ambos lenguajes documentales, hay excepciones que demuestran que un mismo concepto puede estar adscrito a diversas temáticas. En el caso de Eurovoc, términos como *Donación*, *Participación* o *Pobreza* son abordados desde una perspectiva económica y no social, de la misma manera que *Secuestro* está vinculado a una cuestión legal; igualmente, el tesoro europeo entiende por *Radio* el elemento químico radioactivo y no al medio de comunicación. Contrastando los resultados con los del

lenguaje UNESCO, que no presenta este tipo de inexactitudes, se podría decir que este se adscribe mejor a la actividad real del Servicio de Asuntos Sociales.

En cuanto al resto de etiquetas, se encuentran diversas situaciones:

- Hay ciertos términos que aparecen en plural en la lista de etiquetas mientras que en los tesauros lo encontramos en singular. Este es el caso de las etiquetas *Alimentos, Drogas, Empresas, Inmigrantes, Marcas, Materiales, Niños, Niños soldado, Redes Sociales, Refugiados y Tejidos*. Por tanto, unos 11 términos (un 5,88% del total) son similares a los recogidos en las listas controladas.
- Ciertas etiquetas de la lista del Servicio de Asuntos Sociales están compuestas por varias palabras, lo que dificulta la equivalencia en el tesoro; sin embargo, aparecen como conceptos independientes. Tal es el caso, por ejemplo, de *Año Voluntariado Europeo*, una etiqueta en la cual *voluntariado* y *europeo* (sólo en el Eurovoc) son términos recogidos en los lenguajes documentales.
- Un número notable de conceptos, especialmente en el caso de Eurovoc, aparecen de manera recurrente en el tesoro sin tener una entidad propia. Por ejemplo, la etiqueta *ambiental* aparece exclusivamente en el lenguaje controlado en *educación ambiental*. En este mismo sentido, *campañas* aparece en ambos tesauros como *Campañas de* y por tanto, vinculadas a acciones concretas, que nada tiene que ver con el sentido genérico de la etiqueta ni con la actividad real del SAS.
- La ausencia en ambos tesauros de conceptos generales como pueden ser *Discapacidad, Género, Médula, Oncología, Solidaridad, Emprendimiento, Animación*, entre otros ejemplos. Si bien es cierto que no son tesauros centrados en cuestiones médicas, por lo que se entendería la ausencia de ciertos términos, en otros casos, contando con apartados centrados en cuestiones sociales, resulta llamativa la ausencia de conceptos relacionados con el trabajo social.
- La ausencia de palabras de uso frecuente en el lenguaje natural que no suelen estar recogidas en los tesauros generales, pero que pueden ser de utilidad dentro del Servicio de Asuntos Sociales, como es el caso de *Fiesta, Charla, Bocado, Mercadillo, Mayores*, etc.

Teniendo en cuenta todos estos apuntes, se podrían apuntar una serie de observaciones empíricas sobre el uso de los tesauros en el etiquetado del Blog del SAS:

- Según este estudio, sólo un 1 de cada 5 términos utilizados en el etiquetado anterior del Servicio de Asuntos Sociales aparece total o parcialmente en los tesauros Unesco y Eurovoc. Si bien este escaso porcentaje puede venir asociado a la utilización de etiquetas poco adecuadas, podría evidenciar de igual manera una falta de actualización de los tesauros o su escasa adecuación a las necesidades de un recurso web como el del blog del SAS (a destacar el caso de los nombres propios o “identificadores”).

- Si estos tesauros genéricos, disponibles en línea y actualizables presentan ausencias notables de términos, otros lenguajes documentales escritos incluso antes de la utilización generalizada de Internet (me remito, por ejemplo, al clásico “*Tesoro de Servicios Sociales*” del Imsero del año 1986) difícilmente se ajustarán a las necesidades y realidades actuales tanto del Servicio de Asuntos Sociales como de su actividad web.
- Cada tesoro puede utilizar diversas palabras para definir una misma realidad. Por ejemplo, mientras el Eurovoc, al igual que el SAS, opta por la denominación *Cooperación al desarrollo*, UNESCO habla de *Cooperación internacional* o de *Ayuda al desarrollo*. Esto demostraría lo indicado por Gómez Díaz sobre que ante una misma realidad, dos tesauros pueden utilizar diversos conceptos.
- Igualmente, como se ha indicado anteriormente, un mismo término puede estar asociado a diversos microtesauros que no tienen nada que ver con el significado original de la etiqueta, por lo que es importante seleccionar lenguajes documentales acordes a la temática tratada en la medida de lo posible.
- En Internet se puede etiquetar mediante palabras en plural, mientras que en los tesauros no es habitual. Si bien en muchos casos se podría utilizar el singular, no siempre es apropiado: por ejemplo, *Redes Sociales* es una etiqueta habitual y ampliamente aceptada en la Red mientras que en el tesoro aparecerá como *Red Social* (asociado, además, al microtesoro *Sistemas Sociales* de UNESCO, por lo que no tiene nada que ver con la realidad cibernética).
- El tesoro dota en ocasiones a cada término o conjunto de términos de una complejidad y detalle que no siempre es necesaria ni apropiada en Internet. Por ejemplo, el SAS cuenta con la etiqueta *Crisis*, genérica y aplicable a diversos contextos, cuando en los tesauros aparece en todo momento como *crisis política*, *crisis económica*, *crisis internacional* o *crisis energética*, un nivel de detalle que se podría alcanzar en un blog con la utilización de otra etiqueta como Política, Economía, Internacional, Energía, etc.
- En este mismo sentido, la utilización de todos los términos de un tesoro o incluso de un microtesoro en el etiquetado no suele ser necesario, ya que en ocasiones o bien no son aplicables a la realidad descrita, o bien resultan redundantes con respecto a otro término, lo que dificulta la recuperación de información.

5.4.3. La indización automática

Uno de los problemas planteados a la hora de realizar una folksonomía para el Servicio de Asuntos Sociales era el volumen de información a etiquetar y el escaso tiempo que los profesionales disponen para realizar una tarea de estas características. Por este motivo, podría ser interesante en este caso concreto la utilización de sistemas automáticos o semi-automatizados para la selección de los descriptores.

Para ello, tal y como se indicó en el apartado de Metodología, se ha utilizado a la hora de analizar las posibilidades que este método de indización web ofrece el paquete Mallet, generalmente utilizado para la clasificación documental automática pero que igualmente permite conocer los términos más relevantes de los documentos (en este caso, la muestra cuenta con 330 artículos guardados en texto plano), exceptuando aquellos reunidos en una lista de palabras vacías, en la cual se pueden recoger todos aquellos términos que se consideren irrelevantes. Para la selección de estas palabras, se ha utilizado la lista por defecto del extractor terminológico Simple Extractor, creado por la empresa Dail Software, con sede en la Universidad Politécnica de Madrid.

Así, se ha ordenado al programa la extracción de los 300 términos (aunque Mallet sólo extraerá 299 conceptos) con mayor presencia en la totalidad de la muestra, a excepción de las palabras vacías.

En el fichero resultante se ordenan los términos en función de su relevancia o peso. Así, las palabras más utilizadas son aquellas que aparecen en cada uno de los ficheros para diferenciar los diversos apartados que estos contienen (URL, Text, Title), seguido del directorio HTTP y las siglas USAL y SAS, que aparecen en la dirección web. Todos estos conceptos no se pueden considerar descriptivos de la actividad del SAS, por lo que podrían integrarse con posterioridad en la lista de palabras vacías.

Si exceptuamos estos términos, los 10 conceptos más utilizados por los profesionales del SAS a la hora de redactar las entradas del blog serían los siguientes:

TÉRMINOS
VOLUNTARIADO
SALAMANCA
H
DÍA
PERSONAS
SOCIAL
ACTIVIDADES
UNIVERSIDAD
PROGRAMA
CURSO

Tabla 5. Términos más frecuentes del Blog del Servicio de Asuntos Sociales (resultados de Mallet).

A excepción de la letra h, que puede considerarse como errata, el resto de términos han sido recogidos en la anterior lista de etiquetas, bien con entidad propia, bien asociados a otras palabras (“Personas sin hogar”); esto demuestra que no sólo son frecuentes, sino que además podrían tener carácter descriptivo.

Para trabajar con la lista de etiquetas se han ordenado los 299 unitérminos por orden alfabético, tal y como aparecerían en una folksonomía habitual. Dentro de este listado se han identificado conceptos no relevantes.

- 12 ejemplos de erratas, generalmente letras, sílabas (os) y símbolos (*iii*).
- Una serie de conceptos relacionados con el vocabulario web, alguno de ellos anteriormente citado, a lo que habría que sumar otros casos como *org*, *com* o *youtube*. Hay un total de 10 términos de estas características.
- Si bien muchos verbos estaban recogidos en la lista de palabras vacías, se cuenta con 15 conceptos (entre infinitivos, participios y formas conjugadas) asociados a verbos como celebrar, colaborar, dirigir, organizar, participar, promover, de uso habitual en esta institución.
- 14 palabras recogidas en la categorización, por lo que si esta se realiza correctamente, resultaría redundante e innecesaria la utilización de esa etiqueta. Tal es el caso de *voluntariado*, *discapacidad*, *drogas*, etc.
- Es habitual encontrar palabras tanto en plural como en singular. Si bien en algunos casos se podrían utilizar ambas (*Derecho* y *Derechos*), en la mayor parte se podría optar por una u otra. Se encuentran hasta 21 ejemplos.
- Nombres asociados a cuestiones temporales, como es el caso de horas, días, meses y estaciones, con escaso poder descriptivo sobre la actividad. El blog cuenta con un calendario que nos permite localizar los artículos en función de la fecha, por lo que no sería necesaria la introducción de etiquetas de este tipo. 26 términos entrarían dentro de esta categoría.
- Es frecuente que las entradas informen de actividades que requieren inscripción, por lo que términos como *datos*, *nombre*, *correo*, *electrónico* pueden estar vinculados a este tipo de contenido. Se podrían considerar poco relevantes, ya que con catalogar este tipo de entradas en “Participa” (categoría destinada a invitar a la colaboración universitaria en eventos) sería suficiente. 8 términos entran en este grupo.
- 20 palabras que se podrían asociar a espacios físicos en los que se llevan a cabo las actividades realizadas por el SAS. Desde conceptos tan concretos como *Mesa* a otros más generales como *Ciudad*, pasando por otros como *Entrada*, *Edificio*, *Salón*, *Calle*, todos ellos tienen en principio escaso poder descriptivo.

Sin embargo, hay que señalar que dentro de esta categoría entran términos como *Facultad* y *Enfermería*, que podrían ser utilizados a la hora de describir un lugar concreto de uso frecuente, al igual que ocurría en la anterior lista con *Hospedería Fonseca*.

- Se podrían utilizar igualmente etiquetas asociadas a personas si esto se considera relevante, pero de entrada, 5 términos como son *Director*, *Dña*, *José*, *Juan* o *Sánchez*.
- 10 palabras derivadas (adjetivos) de términos recogidos en la lista, con escasa entidad propia aunque podrían estar asociados a otro término. Hablamos, por ejemplo, de la etiqueta *información* e *informativa*, que podría utilizarse en un descriptor como *charla informativa*.
- Por último, se han encontrado 20 palabras de uso frecuente en el lenguaje natural y que no necesariamente se consideran palabras vacías. En este grupo podríamos hablar de conceptos como *Etc*, *Euros* o palabras que pueden formar parte de expresiones como *ámbito*, *forma* o *nivel*.

Tras este análisis, si eliminásemos todos estos términos por ser poco descriptivos, la lista se quedaría en la mitad (137). Sin embargo, si bien en muchos casos, como el de las erratas, es evidente que estos conceptos no deben estar recogidos, no siempre es así. Por ejemplo, *Plaza*, como palabra independiente asociada a un lugar físico, no tiene gran valor descriptivo, pero en la antigua lista de etiquetas aparece asociada a la palabra *Mayor*, haciendo referencia a una asociación de la ciudad de Salamanca, Plaza Mayor, con la que el SAS tiene firmado un acuerdo de colaboración. Por tanto, uno de los problemas que plantea este sistema es que contamos con una lista de unitérminos, que se deben valorar sin conocer el contexto de las entradas, y que de manera individual pueden no ser útiles pero que pueden configurar junto con otra palabra una etiqueta apropiada (por ejemplo, *Comercio Justo*).

Resulta por ello esencial que personas, y en este caso, los profesionales de la institución, realicen un trabajo de asociación de términos y de filtrado, en función de la información que se desea etiquetar y cómo se desee etiquetar, por lo que es importante conocer de primera mano el tipo de información generada. Si ese trabajo se realizase, el uso de programas como Mallet podría ser interesante, ya que conforme los términos se descartan, estos podrían ser incluidos en la lista de palabras vacías y los resultados serían cada vez más apropiados. Por este motivo, el uso de este tipo de programas no puede ser considerado como verdaderamente automático, ya que la labor del profesional determina que los resultados sean unos u otros.

Igualmente, esta muestra con 300 términos debe ampliarse, ya que al priorizar palabras muy utilizadas, se podrían descartar otras que no son frecuentes pero que tienen alto valor descriptivo para una institución como el SAS. En esta lista no aparecen, por ejemplo, nombres de países (sólo aparece *España*) o de asociaciones (como Entreculturas o Feaps) que sí aparecen en la anterior y que son importantes a la hora de describir noticias de Cooperación o de entidades con las que se colabora a nivel

local. De hecho, sólo 81 palabras de las 299 (el 27,1%) aparecen en la antigua lista del Servicio de Asuntos Sociales, que supera los 200 unitérminos.

Contar con palabras de uso frecuente a la hora de redactar las noticias, y por tanto, basarse en el lenguaje natural de los profesionales para describir es otro punto a favor de este sistema, aunque se recomendaría en la medida de lo posible la selección de términos genéricos y la consulta de tesauros, ya que si esto no se realizase, se podrían encontrar numerosas redundancias, ya que el programa no identifica relaciones de sinonimia, polisemia y palabras derivadas. El filtrado de estas duplicidades recaería, por tanto, en manos del profesional.

6. CONCLUSIONES

Acorde a la finalidad de este Trabajo Fin de Máster, donde se han analizado las características de la página web del Servicio de Asuntos Sociales y más concretamente su Blog, se ha considerado oportuno a modo conclusivo apuntar una serie de valoraciones finales que podrían aplicarse en este recurso para la mejora de la recuperación de información y la organización de los contenidos digitales.

La estructura de las categorías en el blog actual del SAS puede ser considerada como correcta, ya que obedece a una lógica interna del Servicio de Asuntos Sociales. Sin embargo, el escaso nivel de categorización unido a los errores de asignación lleva a que se recomiende:

- El establecimiento de una política de categorización que incluya al menos:
 - Una definición exacta de cada una de las categorías y el contenido al que debe hacer referencia.
 - La homogeneidad en cuanto a la asignación de categorías (haciendo especial énfasis en las subcategorías, determinando si estas estarán o no asociadas a su categoría madre).
 - La organización del trabajo grupal para la asignación y la revisión periódica de las categorías.
- La categorización por Áreas de los artículos importados del blog de Diarium, para facilitar así su recuperación.
- La revisión de toda la categorización realizada hasta el momento, debido a los problemas comprobados en ciertos artículos.

En cuanto a la lista de etiquetas anteriormente utilizada por el Servicio de Asuntos Sociales, el inconveniente más frecuente que esta presenta es el alto nivel de duplicidades; se trata de una consecuencia derivada del intermitente interés de los agentes del Servicio por la indización en el antiguo Blog de Diarium. Por ello, se podría concluir que con una correcta planificación similar a la sugerida en el caso de las categorías, que incluya el trabajo grupal a la hora de asignar las etiquetas y la revisión regular de los términos utilizados, se reducirían notablemente los errores cometidos anteriormente.

Igualmente, la revisión de la antigua lista de etiquetas ha permitido la obtención de una serie de apuntes a tener en cuenta a la hora de realizar una nueva folksonomía para el Blog del SAS:

- Que la creación de esta lista de etiquetas se llevó a cabo de forma individual, pero no por ello se puede considerar como una muestra de etiquetado egoísta. Esta modalidad de indización se debe evitar ya que se trata de un recurso público y profesional.

- Que a la hora de realizar una lista de etiquetas es importante tener en cuenta las necesidades propias del Servicio de Asuntos Sociales, por lo que una parte del etiquetado seguirá siendo de corte amiguista.
- Que se debe procurar, en la medida de lo posible, que las etiquetas sean claramente descriptivas para cualquier tipo de usuario (etiquetado altruista).
- Que se debe evitar todo tipo de etiqueta “popular” ya que se trata de una página web de corte profesional en la que el principal objetivo es facilitar la recuperación de la información dentro del propio recurso.

A la hora de plantearse las diversas alternativas metodológicas a la hora de realizar una folksonomía, en los tres niveles de estudio (lenguaje natural, controlado e indización automática), se ha llegado a una serie de conclusiones:

- En cuanto a la modalidad de lenguaje natural, la opción grupal sería la más adecuada y adaptada a la realidad del Servicio de Asuntos Sociales. El hecho de que se trate de un recurso profesional, que involucra a diversos agentes sociales y temáticas lleva a que el etiquetado personal no sea el más recomendable. Al mismo tiempo, la social se descartaría, entre otros motivos, a la limitación tecnológica de las bitácoras, que impiden la asignación de conceptos por parte de los usuarios.

Así, al igual que ocurre en el resto de blogs, la asignación de los conceptos descriptivos debería ser una responsabilidad de los autores de la información, en este caso el personal del SAS. Ellos son los que conocen la temática y el ámbito de acción de todo el Servicio y de cada una de sus unidades, por lo que igualmente tienen la capacidad de asignar los términos apropiados para la descripción de los contenidos.

Sin embargo, se ha apuntado dos problemas de este tipo de modalidad en el caso del Servicio de Asuntos Sociales: que se puedan cometer con el paso del tiempo errores similares a los de la anterior folksonomía, y la falta de tiempo a la hora de asignar los descriptores.

- Los lenguajes documentales (en este caso concreto, el Tesouro UNESCO y Eurovoc) no se plantean como una solución adecuada a los intereses de un recurso digital. Los tesauros no han sido recursos creados ad hoc para la elaboración de listas de etiquetas, y la rigurosidad y el detalle de los lenguajes documentales (lo que exige tiempo para desarrollar ese trabajo intelectual) entra en contraposición con la simplicidad y la actualización constante de las etiquetas basadas en el lenguaje natural.

El problema más evidente del uso de lenguajes controlados en la ausencia de nombres propios, lo que impide la utilización de identificadores que permiten y detallar actividades concretas, una funcionalidad especialmente importante en el caso del SAS (54 etiquetas de las 193 entrarían dentro de este grupo en la anterior folksonomía).

Sin embargo, la consulta de tesauros podría paliar uno de los problemas expuestos por el personal del Servicio como es la inecuación de ciertas etiquetas. Así, se recomienda la consulta de lenguajes documentales temáticos y ciertos microtesauros para la selección de términos correctos.

- La indización automática, para la cual se ha utilizado el clasificador documental Mallet, no se recomendaría tampoco a día de hoy. Aunque el personal conociese el funcionamiento de programas de este tipo, lo que actualmente no es así porque su orientación laboral nada tiene que ver con el campo de la Documentación, lo cierto es que este método no reduce notablemente la carga de tiempo destinado a la realización de una folksonomía, ya que hay que crear archivos con entradas y palabras vacías, y realizar el filtrado y asociación de unitérminos.

Al mismo tiempo, si la responsabilidad de la lista de etiquetas recayese en terceras personas o personal que no conozca los términos de otras unidades, esta metodología sería interesante ya que permitiría conocer los términos utilizados por los profesionales de forma automática; sin embargo, si el personal se ocupase de esta tarea, no habría una diferencia sustancial entre el uso de este sistema y el de unas etiquetas asignadas por ellos con una correcta revisión periódica.

En definitiva, ningún método por si solo daría una solución real a la hora de asignar términos para el blog del Servicio de Asuntos Sociales, y cabría pensar en una opción que tenga en cuenta todas las modalidades. Siguiendo el modelo de la anterior folksonomía, que está lejos de ser un recurso totalmente descartable, se podría realizar un trabajo común de asignación y revisión teniendo en cuenta la consulta de tesauros ante dudas y la detección automática de términos para conocer los conceptos más habituales. Sin duda esta labor requiere tiempo, pero si se realizase una adecuada división del trabajo y todos los agentes estuviesen igualmente involucrados en la tarea, se podría obtener un resultado verdaderamente profesional y adecuado al recurso.

Este trabajo abre otras líneas de investigación aplicada, como pueda ser la comparación de la interoperabilidad semántica de los lenguajes documentales mencionados, usando herramientas como podrían ser los mapas conceptuales. También sería oportuno comparar la organización de contenido que se da en el SAS con otras unidades de temática semejante, sean o no universitarias. Por último, otra línea de investigación posible sería poder analizar el comportamiento en las búsquedas de usuarios reales que interactúan con los recursos informativos del SAS.

7. BIBLIOGRAFÍA

ATTALI, J. (1995). Les labyrinthes de l'information. *Le Monde* (9 noviembre 1995). Disponible en: <http://www.attali.com/actualite/blog/nouvelle-economie/les-labyrinthes-de-linformation> [Fecha de consulta: 10 julio 2015].

BURGOS BORDONAU, E y PALACIOS GÓMEZ, J.L. (2009). Estudio bibliométrico y de calidad de la revista Cuadernos de Trabajo Social (1987-2008). *Cuadernos de Trabajo Social*, 83. Vol. 22 (2009) p. 83-108. Disponible en: <http://revistas.ucm.es/index.php/CUTS/article/viewFile/CUTS0909110083A/7502> [Fecha de consulta: 18 agosto 2015]

CASTILLO BLANCO, L. (2006). Elaboración de un tesoro de información de actualidad y conversión en red semántica para su empleo en un sistema de recuperación periodístico. Tesis Doctoral. Universitat de València. Departament d'Història de la Ciència i la Documentació. Disponible en: <http://hdl.handle.net/10803/9982> [Fecha de consulta: 2 julio 2015].

CÓMO usar el Planificador de palabras clave para obtener ideas para palabras clave y previsiones de tráfico. *Ayuda de AdWords*. [en línea]. Disponible en: <https://support.google.com/adwords/answer/2999770?hl=es>. [Fecha de consulta: 17 julio 2015].

DÁVILA PESSAGNO, M. I. (2005), Tesoro de Trabajo Social. Rojas Eberhard. 188 p. ISBN: 9589121772.

ESCOLANO, D. (2013). Adiós a la herramienta de palabras clave de Adwords. *Human Level Communications*. [en línea]. (22 agosto 2013). Disponible en: <http://www.humanlevel.com/articulos/google-adwords/adios-a-la-herramienta-de-palabras-clave-de-adwords.html> [Fecha de consulta: 17 julio 2015].

EUROVOC, tesoro multilingüe de la Unión Europea (en español). [en línea]. Disponible en: <http://eurovoc.europa.eu/drupal/?q=es> [Fecha de consulta: 15 agosto 2015].

EXTRACTOR terminológico: "SimpleExtractor" V1.1.2. *Dail Software. Tienda online*. [en línea]. Disponible en: <http://www.dail-software.com/shop/es/9-extractor-terminol%C3%B3gico-simpleextractor-v112.html> [Fecha de consulta: 20 agosto 2015].

GÓMEZ-DÍAZ (2012). Etiquetar en la Web Social. Editorial UOC. (El profesional de la información: 9). 114 p. ISBN: 9788490290132.

GONZÁLEZ, D. y MARCOS, M. C. (2013). Responsive web design: diseño multidispositivo para mejorar la experiencia de usuario. *BiD: textos universitaris de biblioteconomia i documentació*, Diciembre, núm. 31. Disponible en: <http://bid.ub.edu/es/31/gonzalez2.htm> [Fecha de consulta: 20 julio 2015].

GONZALO PENELA, C. (2004). La selección de palabras clave para el posicionamiento en buscadores. *Hipertext.net* [en línea] núm. 2, 2004. Disponible en: http://www.upf.edu/hipertextnet/numero-2/palabras_clave.html [Fecha de consulta: 15 julio 2015].

GORMAN, M. (1995). The corruption of cataloging. *Library Journal* (15 septiembre, 1995), p. 32-34. Disponible en: <http://murraylib630.org/GormanCorruptionCataloging.pdf> [Fecha de consulta: 10 julio 2015].

HASSAN MONTERO, Y. (2006). Indización Social y Recuperación de Información. *NSU: No sólo usabilidad: Revista sobre personas, diseño y tecnología* [en línea]. (16 noviembre 2006). Disponible en: http://www.nosolousabilidad.com/articulos/indizacion_social.htm [Fecha de consulta: 18 julio 2015].

HISTORIA de los buscadores (2011). *Historia de la Informática* [en línea], (11 enero 2011). Universitat Politècnica de València. Disponible en: <http://histinf.blogs.upv.es/2011/01/11/1808/> [Fecha de consulta: 15 julio 2015].

JUÁREZ, M. (ed.) (1993). Trabajo social e investigación: temas y perspectivas. UPCO. 157 p.

LAGUENS GARCÍA, J.L. (2006). Tesoros y Lenguajes Controlados en Internet. *Anales de Documentación*, núm 9, p. 105-121. Disponible en: <http://revistas.um.es/analesdoc/article/viewFile/1391/1441> [Fecha de consulta: 12 julio 2015].

LAMARCA LAPUENTE, M.J. (2006). Indización Automática. Hipertexto: el nuevo concepto de documento en la cultura de la imagen. Tesis Doctoral. [en línea]. Disponible en: http://www.hipertexto.info/documentos/indiz_automat.htm [Fecha de consulta: 21 julio 2015].

LANCASTER, F.W. (2002). El control del vocabulario en la recuperación de información. Universitat de València. 286 p. ISBN: 8437054443.

MALLET Machine Learning for Language Toolkit. [en línea]. Disponible en: <http://mallet.cs.umass.edu/> [Fecha de consulta: 20 agosto 2015].

MARGAIX ARNAL, D. (2008). Informe APEI sobre web social. Asociación Profesional de Especialistas en Información. 64 p. ISBN: 9788469177266. Disponible en: <http://eprints.rclis.org/12506/1/informeapeiwebsocial.pdf> [Fecha de consulta: 15 agosto 2015].

MATAMOROS SEGURA, L y RAMÍREZ MORA, L. (2001), Tesouro en ciencias sociales. Centro de Información Documental de la Facultad de Ciencias Sociales. 304 p. ISBN: 9968993816.

MÉNDEZ RODRIGUEZ, E.M y MOREIRO GONZÁLEZ, J.A. (1999). Lenguaje natural e Indización Automática. *Ciencias de la Información*, vol. 30, nº 3, septiembre 1999, p.11-24. Disponible en: <http://eprints.rclis.org/12685/1/indizacion99.pdf> [Fecha de consulta: 18 julio 2015].

MOREIRO GONZÁLEZ, J. A. (2007). La representación de los contenidos digitales: de los tesauros automáticos a las folksonomías. *VI Workshop CALSI: Información digital: nuevas perspectivas en la sociedad del conocimiento* (Valencia 14, 15 y 16 de noviembre de 2007). Disponible en: <http://www.terras.edu.ar/aula/tecnicatura/16/biblio/16folksonomias.pdf> [Fecha de consulta: 20 mayo 2015].

NO estás posicionando: El meta tag Keywords es obsoleto desde 2009 (2013). *Técnica Seo* [en línea]. (1 septiembre 2013). Disponible en: <http://www.tecnicaseo.com/seo/tecnicas/meta-tag-keywords-posicionamiento> [Fecha de consulta: 30 agosto 2015].

PÉREZ SANCHIDRIÁN, E., CAMPOS POSADA, R. y CAMPOS POSADA, G.E. (2014). Etiquetado social: un modelo de representación de la información en la blogosfera. *Biblios* [en línea] University of Pittsburgh. Núm. 56. Disponible en: <http://biblios.pitt.edu/ojs/index.php/biblios/article/viewFile/187/207> [Fecha de consulta: 17 julio 2015].

PINTO MOLINA, M. (1993) *Análisis Documental. Fundamentos y procedimientos*. Editorial Eudema Universidad, 270 p., (2ª ed.). ISBN: 8477540705.

PLANIFICADOR de Palabras Clave. Google AdWords. Disponible en: <https://adwords.google.es/KeywordPlanner> [Fecha de consulta: 13 julio 2015].

POR las nubes. Blog sobre documentación, sistemas de representación, nubes de etiquetas... [en línea]. Disponible en: <https://porlasnubes.wordpress.com/> [Fecha de consulta: 17 julio 2015].

RIVERA SÁNCHEZ, C. (2013). De las folksonomías o la Torre de Babel. *Infotecarios* [en línea]. (30 enero 2013). Disponible en: <http://www.infotecarios.com/de-las-folksonomias-o-la-torre-de-babel/> [Fecha de consulta: 20 julio 2015]

RODRÍGUEZ ARMESTO, B. (2013). Cómo y por qué planificar las etiquetas de un blog. *40 de Fiebre* [en línea] (4 junio 2013). Disponible en: <http://www.40defiebre.com/como-planificar-etiquetas-blog/> [Fecha de consulta: 2 agosto 2015].

RODRÍGUEZ PARADA, C. (2007). Los catálogos e inventarios en la historia del libro y de las bibliotecas. *BiD: textos universitaris de biblioteconomia i documentació*, junio, núm. 18. Disponible en: <http://bid.ub.edu/18rodri4.htm> [Fecha de consulta: 11 julio 2015].

RODRÍGUEZ YUNTA, L. (2009). Etiquetado libre frente a lenguajes documentales. Aportaciones en el ámbito de Biblioteconomía y Documentación. *IX Congreso ISKO-España*, (Valencia 11-13 Marzo 2009). Nuevas perspectivas para la difusión y organización del conocimiento. Actas del Congreso. Valencia: UPV, vol 2, pp. 832-845 Disponible en: <http://eprints.rclis.org/12571/> [Fecha de consulta: 25 mayo 2015].

SERVICIO de Asuntos Sociales de la Universidad de Salamanca. Página web oficial. Disponible en: <http://sas.usal.es/>. [Fecha de consulta: 5 mayo 2015].

TAGS, etiquetas o nube de palabras en el SEO: ventajas y desventajas. (2015). *SEM, SEO y mucho +* [en línea] (22 abril 2015). Disponible en: <http://blog.semseoymas.com/tags-etiquetas-o-nube-de-palabras-en-el-seo-ventajas-y-desventajas-41.htm> [Fecha de consulta: 17 julio 2015].

TESAURO de la Discapacidad (2009). Observatorio Estatal de la Discapacidad. 373 p. Disponible en: <http://observatoriodeladiscapacidad.info/attachments/article/18/tesauro%20de%20la%20discapacidad.pdf> [Fecha de consulta: 17 agosto 2015].

TESAURO de la UNESCO (en español). [en línea]. Disponible en: <http://databases.unesco.org/thesp/> [Fecha de consulta: 15 agosto 2015].

TESAURO de Servicios Sociales. (1986). Instituto Nacional de Servicios Sociales, Secretaria General para la Seguridad Social, Ministerio de Trabajo y Seguridad Social. 146 p.

TESAUROS y Glosarios IEDCYT en Línea. Instituto de Estudios Documentales sobre Ciencia y Tecnología (antes CINDOC). Consejo Superior de Investigaciones Científicas. Disponible en: http://thes.cindoc.csic.es/index_esp.php [Fecha de consulta: 17 agosto 2015].

USO excesivo de palabras clave. *Ayuda de Search Console* [en línea]. Disponible en: <https://support.google.com/webmasters/answer/66358?hl=es> [Fecha de consulta: 17 julio 2015].

VANDER WAL, T. (2005). Explaining and Showing Broad and Narrow Folksonomies. *Walderwal.net* [en línea] (21 febrero 2005). Disponible en: <http://www.vanderwal.net/random/entrysel.php?blog=1635> [Fecha de consulta: 20 julio 2015].

8. APÉNDICES

8.1. Lista de etiquetas original del Blog de Diarium del Servicio de Asuntos Sociales de la Universidad de Salamanca

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
#BRINGBACKOURGIRLS	1
50 ANIVERSARIO	1
ABRAZO	1
ACAPARAMIENTO DE TIERRAS	2
ACTIVIDADES LÚDICAS	1
ALIMENTOS	1
ALLEGRO	1
AMAZONIA	1
AMBIENTAL	17
AMNISTÍA INTERNACIONAL	2
ANIMACIÓN INFANTIL	1
AÑO VOLUNTARIADO EUROPEO	1
APOYO ESCOLAR	1
ARÉVALO	1
ARGUMENTOS	1
ASDECOBA	1
ASOCIACIÓN VIENTO NORTE SUR	1
ASPROCES	1
AYUDA HUMANITARIA	2
AYUNTAMIENTO	1
BOCADILLO	1
BOLETÍN INJUVE	2
CAFÉ TERTULIA	2
CAMBIO SOCIAL	1
CAMBOYA	1
CAMPAMENTO URBANO	1

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
CAMPAÑA	1
CAMPOS DE TRABJO	1
CAMPUS INCLUSIVOS	1
CAMPUS SIN LÍMITES	1
CÁRITAS	6
CASTILLA Y LEÓN	2
CATÁSTROFE	1
CGT	1
CHARLA	2
CIBERVOLUNTARIOS	1
CICLO DE CINE	1
CIES	1
CINEFORUM	2
CÍRCULOS DE SILENCIO	4
CLIPMETRAJES	1
COLTÁN	1
COMERCIO JUSTO	4
COMUNICACIÓN	1
CONCENTRACIÓN	1
CONCIERTO SOLIDARIO	1
CONCURSO	1
CONCURSO DE CORTOS	1
CONFERENCIA	1
CONGD	1
CONSUMO RESPONSABLE	2
COOPERACI	0
COOPERACIÓN	1
COOPERACIÓN AL DESARROLLO	4
COORDINADORA DE ONGD	1
COORDINADORA ONG PARA EL DESARROLLO	1

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
CORAZÓN	1
CRISIS	2
CUMPLEAÑOS	1
CURSO	2
CURSOS	2
DECÁLOGO	1
DECÑALOGO	0
DEJANDO HUELLA	0
DERECHO INDÍGENAS	1
DERECHOS	1
DERECHOS HUMANOS	2
DESIERTO	2
DÍA	1
DIÁLOGOS CROADOS	0
DISCAPACIDAD	2
DISCRIMINACIÓN	1
DONACIÓN	2
DROGAS	1
EL SAVOR	1
EMPLEABILIDAD	1
EMPLEO	1
EMPREDIMIENTO CULTURAL	1
EMPREDIMIENTO SOCIAL	1
EMPRESAS	1
ENTRECULTURAS	2
FEAPS	1
FELICIDAD	1
FÉLIX LÓPEZ	1
FERIA	1
FIESTA	1

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
FILIPINAS	1
FIRMAS	1
FLASHMOB	1
FORMACIÓN	1
FORMACIÓN BÁSICA	1
FUNDACIÓN SECRETARIADO GITANO	1
FUNDACIÓN TORMES	1
FUNDACIÓN TORMES - EB	1
GALA BENÉFICA	1
GÉNERO	2
GREENPEACE	1
HEMODONACIÓN	1
HOSPEDERÍA FONSECA	1
INFANCIA	2
INMIGRACIÓN	2
INMIGRANTES	1
INNOVACIÓN SOCIAL	1
INSTITUTO IBEROAMERICANO	1
INTERCULTURALIDAD	1
INTERGENERACIONAL	2
INTERMÓN OXFAM	5
INTERNACIONAL	1
JORNADAS	6
JUSTICIA SOCIAL	1
JUVENTUD	4
KAIKIT	1
LA PARTIOLA	1
LANA	1
MANOS UNIDAS	3
MANOS UNIDOS	1

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
MARCAS	1
MARRUECOS	1
MATERIALES	0
MAYORES	1
MEDIO AMBIENTE	5
MÉDULA	1
MENORES	0
MERCADILLO	1
MALAKAI (ERRATA)	1
MUNDO CATARATA	1
NIGERIA	1
NIÑOS	1
NIÑOS CON CÁNCER	1
NIÑOS SOLDADOS	1
NOTICIAS	0
OFICINA VERDE	4
ON-LINE	1
ONCOLOGÍA	1
OPERACIÓN BOCATA	1
ÓRGANOS	1
OXFAM INTERMON	8
PÁNDANO	1
PANDORA	2
PARAGUAY	1
PARTICIPACIÓN	2
PERSONAS SIN HOGAR	2
PIZARRALES	1
PLAZA MAYOR	1
POBREZA	1
PROGRAMACIÓN	1

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
PROGRESO	1
PROMOCIÓN Y DESARROLLO SOCIAL	1
PVE	0
PYFANO	2
RACISMO	1
RADIO	1
RECICLAJE	2
RED DE VOLUNTARIADO SOCIAL	1
RED SALAMANC	0
RED SALAMANCA	2
RED ZAMORA	0
REDES SOCIALES	1
REFUGIADOS	1
RETOS GLOBALES	1
RISOTERAPIA	2
SALAMANCA	2
SALAMANCA ABRIGA	1
SALAMANCA ACOGE	1
SALUD	2
SANGRE	1
SAS	2
SECUESTRO	1
SEMANA VERDE	1
SENSIBILIZACIÓN	2
SIPPE	1
SIRIA	1
SOCIOSANITARIA	1
SOLIDARIDAD	3
SÚMATE	0
SVE/EVS/VOLUNTARIADO EUROPEO	0

ETIQUETAS	Nº MENCIONES BLOG DIARIUM
TALLER	1
]	1
TEJER	1
TEJIDOS	1
TELEFONO DE LA ESPERANZA	1
TERCERA EDAD	1
TERTULIA CAFÉ	1
TIFÓN HAIYAN	1
TRANSPLANTE	1
TU PUNTO DE MIRA	1
TURISMO RESPONSABLE	1
UNDIAPARADAR	1
UNIVERSIDAD	1
VALLADOLID	1
VIAJE	1
VIAJES SOLIDARIOS	1
VOLUNTARIADO	8

8.2. Muestra de los 299 unitérminos extraídos con Mallet

POR ORDEN ALFABÉTICO	POR RELEVANCIA
ABRIL	URL
ACCION	TEXT
ACTIVIDAD	TITLE
ACTIVIDADES	USAL
ACTO	HTTP
ACTOS	SAS
AMBIENTAL	VOLUNTARIADO
AMBITO	SALAMANCA
ANIMACION	H
ANO	DIA
ANOS	PERSONAS
APOYO	SOCIAL
AS	ACTIVIDADES
ASISTENCIA	UNIVERSIDAD
ASOCIACION	PROGRAMA
ASOCIACIONES	CURSO
ASUNTOS	SERVICIO
AULA	HORAS
AVILA	INFORMACION
AVIVA	PLAZA
AYUDA	DESARROLLO
BENEFICIO	PARTICIPAR
C	VOLUNTARIOS
CABO	CAMPANA
CAFE	SOCIALES
CALLE	INTERNACIONAL
CAMBIO	ANO
CAMPANA	MAYO

POR ORDEN ALFABÉTICO	POR RELEVANCIA
CAMPUS	ASOCIACION
CARITAS	FUNDACION
CARRERA	TU
CASTILLA	PARTICIPACION
CELEBRA	SEMANA
CELEBRACION	DERECHOS
CELEBRARA	DISCAPACIDAD
CENTRO	FACULTAD
CHARLA	PROYECTO
CICLO	JOVENES
CINE	FORMACION
CIUDAD	NOVIEMBRE
CIUDADANIA	EDUCACION
COLABORACION	CAMPUS
COLABORAR	MARZO
COLEGIO	CENTRO
COM	INSCRIPCION
COMERCIO	EUROPEO
COMPROMISO	ABRIL
COMUNICACION	JUEVES
COMUNIDAD	MUNDO
CONCIERTO	LEON
CONCURSO	COOPERACION
CONFERENCIA	ESCUELA
CONSUMO	WWW
CONTACTO	PROYECTOS
CONVOCATORIA	SOLIDARIO
COOPERACION	JORNADAS
CORREO	ESPANA

POR ORDEN ALFABÉTICO	POR RELEVANCIA
CRISIS	COLABORACION
CULTURA	DIAS
CULTURAL	ASUNTOS
CURSO	VIERNES
D	CASTILLA
DATOS	ACCION
DEPORTE	C
DERECHO	PUEDES
DERECHOS	INTERGENERACIONAL
DESARROLLO	JORNADA
DIA	ZAMORA
DIAS	ANOS
DICIEMBRE	CONCURSO
DIRECTOR	OXFAM
DIRIGIDO	INTERMON
DISCAPACIDAD	FEBRERO
DIVERSAS	PODEIS
DNA	DROGAS
DOMINGO	HUMANOS
DONACION	UNIVERSITARIA
DONANTES	TRABAJO
DROGAS	VOLUNTARIO
EDICION	D
EDIFICIO	TE
EDUCACION	ESTUDIANTES
ELECTRONICO	NINOS
EMPLEO	TEATRO
EMPRENDIMIENTO	COM
ENCUENTRO	GRUPO

POR ORDEN ALFABÉTICO	POR RELEVANCIA
ENERO	APOYO
ENFERMERIA	WEB
ENTIDADES	OBJETIVO
ENTRADA	VOLUNTARIA
ENTRADAS	FORMA
ESCUELA	CORREO
ESPACIO	JUVENTUD
ESPANA	ENCUENTRO
ESPANOLA	OCTUBRE
ESPECIAL	COMUNIDAD
ESTUDIANTES	TALLER
ETC	SOLIDARIA
EUROPEA	CINE
EUROPEO	SAN
EUROS	SABADO
EXPERIENCIA	ORGANIZA
EXPERIENCIAS	OS
EXPOSICION	YOUTUBE
FACULTAD	ORG
FAVOR	COMUNICACION
FEBRERO	ACTIVIDAD
FECHA	TIENES
FIESTA	POBREZA
FONSECA	CONVOCATORIA
FORMA	SALUD
FORMACION	VIDA
FORMULARIO	I
FORUM	TALLERES
FUNDACION	ENTRADA

POR ORDEN ALFABÉTICO	POR RELEVANCIA
FUTURO	LUNES
GENERAL	JUNIO
GENERO	CIUDAD
GRATUITA	TORMES
GRUPO	ACTOS
GRUPOS	ENTIDADES
H	ELECTRONICO
HORA	EDICION
HORARIO	CELEBRA
HORAS	MAYORES
HTTP	UNIVERSITARIO
HUMANOS	CHARLA
I	PAIS
IGUALDAD	MIERCOLES
II	CONCIERTO
III	CULTURAL
INFORMACION	AMBIENTAL
INFORMATIVA	CARITAS
INICIATIVA	CALLE
INNOVACION	SOCIEDAD
INSCRIPCION	SALA
INTERCULTURAL	SALON
INTERGENERACIONAL	CULTURA
INTERMON	VERDE
INTERNACIONAL	AYUDA
JORNADA	VOLUNTARIAS
JORNADAS	SANCHEZ
JOSE	DONACION
JOVENES	JULIO

POR ORDEN ALFABÉTICO	POR RELEVANCIA
JUAN	HORARIO
JUEVES	POBLACION
JULIO	SANGRE
JUNIO	V
JUNTA	CELEBRARA
JUSTO	OFICINA
JUVENIL	EXPERIENCIA
JUVENTUD	ASOCIACIONES
LEMA	DICIEMBRE
LEON	DERECHO
LIBRE	MESA
LOCAL	UNIDAS
LUCHA	REALIZARA
LUNES	ORGANIZADO
MADRID	SOLIDARIDAD
MANOS	EXPOSICION
MARCHA	LIBRE
MARTES	EUROS
MARZO	TELEFONO
MAYO	PAISES
MAYORES	AVILA
MEDIOS	ONG
MEDULA	TITULO
MES	PAGINA
MESA	PREMIO
MESAS	MUNDIAL
MESES	LOCAL
MIERCOLES	AULA
MOTIVO	QUIERES

POR ORDEN ALFABÉTICO	POR RELEVANCIA
MUJER	PRETENDE
MUJERES	GENERAL
MUNDIAL	CONTACTO
MUNDO	SENSIBILIZACION
MUSICA	ESPACIO
N	ORGANIZACION
NACIONAL	CABO
NINOS	PUNTO
NIVEL	MES
NOMBRE	PARTICIPANTES
NOVIEMBRE	RECURSOS
NUMERO	SUPERIOR
OBJETIVO	EMPLEO
OBJETIVOS	REALIDAD
OCTUBRE	MUSICA
OFICINA	LEMA
ONCE	PLATAFORMA
ONG	MARTES
ORG	TIEMPO
ORGANIZA	SOMOS
ORGANIZACION	JUAN
ORGANIZACIONES	PREMIOS
ORGANIZADO	MADRID
OS	SITUACION
OXFAM	ORGANIZACIONES
PAGINA	DOMINGO
PAIS	PRESENTACION
PAISES	FORMULARIO
PARTICIPACION	RED

POR ORDEN ALFABÉTICO	POR RELEVANCIA
PARTICIPANTES	DNA
PARTICIPAR	FAVOR
PASOS	MOTIVO
PAZ	CAFE
PERSONA	IGUALDAD
PERSONAS	INICIATIVA
PLATAFORMA	NACIONAL
PLAZA	ESPANOLA
PLAZAS	LUCHA
PLAZO	CAMBIO
POBLACION	CICLO
POBREZA	FORUM
PODEIS	AMBITO
PREMIO	PLAZO
PREMIOS	MESAS
PRESENTACION	EDIFICIO
PRETENDE	MANOS
PREVENCION	MESES
PROGRAMA	PERSONA
PROMOCION	DEPORTE
PROMOVER	DIRECTOR
PROPUESTAS	JUSTO
PROVINCIA	CIUDADANIA
PROYECTO	CRISIS
PROYECTOS	OBJETIVOS
PUEDES	JOSE
PUNTO	FONSECA
PYFANO	INTERCULTURAL
QUIERES	AS

POR ORDEN ALFABÉTICO	POR RELEVANCIA
RADIO	S
REALIDAD	MEDULA
REALIZARA	VIH
RECONOCIMIENTO	PYFANO
RECURSOS	FECHA
RED	MUJERES
REFLEXION	DIRIGIDO
S	PASOS
SABADO	CARRERA
SALA	MARCHA
SALAMANCA	PROVINCIA
SALON	DATOS
SALUD	CELEBRACION
SAN	VERANO
SANCHEZ	DIVERSAS
SANGRE	ETC
SANTA	FIESTA
SAS	COMPROMISO
SEMANA	II
SENSIBILIZACION	ENERO
SERVICIO	COLABORAR
SERVICIOS	SILENCIO
SILENCIO	HORA
SITUACION	COMERCIO
SOCIAL	CONFERENCIA
SOCIALES	AVIVA
SOCIEDAD	COLEGIO
SOLIDARIA	NIVEL
SOLIDARIDAD	GRATUITA

POR ORDEN ALFABÉTICO	POR RELEVANCIA
SOLIDARIO	SVE
SOMOS	EMPRENDIMIENTO
SUPERIOR	ANIMACION
SVE	PREVENCION
TALLER	MUJER
TALLERES	SANTA
TE	GENERO
TEATRO	REFLEXION
TELEFONO	PAZ
TEXT	ASISTENCIA
TIEMPO	BENEFICIO
TIENES	ACTO
TITLE	MEDIOS
TITULO	EXPERIENCIAS
TORMES	PROMOVER
TRABAJO	RECONOCIMIENTO
TU	UNIVERSITARIOS
UNIDAS	N
UNIVERSIDAD	ONCE
UNIVERSITARIA	ENFERMERÍA
UNIVERSITARIO	DONANTES
UNIVERSITARIOS	INNOVACIÓN
URL	RADIO
USAL	FUTURO
V	CONSUMO
VERANO	NUMERO
VERDE	PLAZAS
VIDA	PROMOCIÓN
VIERNES	ENTRADAS

POR ORDEN ALFABÉTICO	POR RELEVANCIA
VIH	ESPECIAL
VOLUNTARIA	NOMBRE
VOLUNTARIADO	INFORMATIVA
VOLUNTARIAS	JUNTA
VOLUNTARIO	SERVICIOS
VOLUNTARIOS	III
WEB	GRUPOS
WWW	EUROPEA
YOUTUBE	PROPUESTAS
ZAMORA	JUVENIL