

---

# Organización del conocimiento para la documentación en periodismo: situación y prospectiva

*Knowledge organization for journalism documentation: state of the art and prospective*

---

**Antonio GARCÍA JIMÉNEZ**

Facultad Ciencias de la Comunicación, Universidad Rey Juan Carlos, C/ Camino del Molino, s/n, Campus de Fuenlabrada, 28943 Fuenlabrada, Madrid, España; antonio.garcia@urjc.es

## Resumen

Se pretende establecer el estado actual, por un lado, de instrumentos como tesauros, ontologías y folksonomías y, por otro, de los procedimientos y modelos de formalización para el análisis y etiquetado semántico, al objeto de determinar los aspectos más relevantes en su aplicación en el dominio periodístico. Tras esta revisión, y tomando en consideración los cambios que se vienen produciendo en el periodismo, se plantean los puntos sobre los que debe girar futuras reflexiones sobre la organización del conocimiento en conexión con la web semántica, los datos abiertos, las necesidades de información y los nuevos modelos de recuperación de información.

**Palabras clave:** Organización del conocimiento. Tesauros. Ontologías. Folksonomías. Marcado semántico. Documentación. Periodismo.

## 1. Introducción

El objetivo de este trabajo es doble. Por un lado, precisar la realidad actual de las técnicas e instrumentos que afectan a la organización y representación del conocimiento, en especial en lo que tiene que ver como los tesauros, las folksonomías y los modelos de análisis y marcado semántico. Por otro lado, delimitar los aspectos concretos vinculados con los vocabularios y los procesos de análisis semántico en los que hay profundizar para poder evaluar la realidad de la organización del conocimiento para la Documentación periodística, partiendo de la premisa de que se trata de un dominio que ha sufrido profundas transformaciones en los últimos años.

La presencia definitiva de lo digital, el cambio en el trabajo de los periodistas y en su relación con las noticias y las fuentes, la creciente tecnificación, el asentamiento definitivo de las redacciones digitales, el disperso uso de la web semántica, así como la relevancia de los medios sociales son fenómenos que dibujan un escenario diferente y que hacen necesaria una revisión de lo que atañe a la organización y representación del conocimiento en este ámbito.

## Abstract

This paper aims to establish the state of the art, on the one hand, about instruments such as thesauri, ontologies and folksonomies and, secondly, procedures and models of formalization for analysis and semantic tagging in order to determine the most relevant aspects in its application in the journalistic domain. Following this review, and taking into consideration the changes that have taken place in journalism, the study points as future reflections focuses: knowledge organization in connection with semantic web, open data, information needs and new models of information retrieval are considered.

**Keywords:** Knowledge organization. Thesauri. Ontologies. Folksonomies. Semantic markup. Documentation. Journalism.

En primer lugar, se aborda el contexto comunicativo en el que nos situamos, caracterizado por fenómenos como el *open data* y el *big data*, especialmente en los elementos que tienen especial incidencia en el plano documental, en general, y en la organización del conocimiento, en particular. A continuación, se hace una revisión de la literatura existente, por un lado, de los lenguajes controlados y, por otro, de los procedimientos de análisis. En tercer lugar, se revisan experiencias aplicadas y académicas ya centradas en el periodismo escrito. Finalmente, se plantean los aspectos centrales necesarios para avanzar en el conocimiento sobre los instrumentos y procedimientos de organización del conocimiento que se aplican y el modo en que se lleva a cabo.

## 2. El contexto comunicativo y periodístico

Es cada vez más patente el impacto que tienen en el periodismo los nuevos avances y desarrollos derivados del *open data* y del *big data*. Como es sabido, con los datos en abierto se busca la eliminación de las restricciones existentes, en di-

ferentes niveles, en la difusión, acceso y reutilización de los datos y la información. Esta realidad impregna diferentes áreas tanto del ejercicio periodístico como de la realidad documental. En gran medida, esta corriente es deudora del trabajo desplegado por los activistas que defienden el libre acceso al conocimiento y está en sintonía con acciones, incluso con elementos estéticos, *hacker*. Por otra parte, la incorporación paulatina del trabajo con algoritmos que ayuden a la interpretación de la gran cantidad de información existente en la red también es uno de los aspectos clave para el análisis de la actividad documental, en general, y de la organización y representación del conocimiento, en particular.

A juicio de Sugimoto, Ding y Thelwall (2012), la propia Documentación no ha sido ajena a los inicios del *big data*, por ejemplo a partir de las diferentes métricas aplicadas al estudio de la web, en concreto en lo que se refiere a la recopilación y análisis de datos web a gran escala, junto a los programas y proyectos de análisis de enlaces. Además, esto se ha hecho en conexión con las fórmulas de integración de la web semántica y los *linked open data*, junto a las herramientas, técnicas y teorías bibliométricas, profusamente empleadas.

Ya en el plano mediático, conforme al trabajo de Stone (2014), es cada vez mayor el interés aplicado de los medios de comunicación, como CCN, Financial Times o The Huffington Post, por el *big data*. En primer lugar, destacan los formatos de visualización. Por otra parte, se presentan como referentes los modelos de análisis de titulares y de adaptación a las audiencias (a través del análisis de comentarios, de la calidad de los productos periodísticos, de métricas *ad hoc*, de preferencias de contenidos, o de comportamientos de navegación, etc.), así como la personalización y segmentación de contenidos periodísticos, el seguimiento de noticias en tiempo real (mediante la identificación, clasificación y alerta de información relevante a usuarios concretos) y, finalmente, la generación de noticias por medio de la condensación de datos y la posterior presentación en formatos concretos.

Una de las vertientes más destacables que muestra conexiones con el hecho documental es el periodismo de datos. En este sentido, Ferreras Rodríguez (2012) caracteriza al periodista de datos alrededor de cuatro ideas, con claras concomitancias documentales: por un lado, la recopilación de datos e información, teniendo en cuenta tanto los datos en abierto como las técnicas de recuperación más avanzadas. En segundo lugar, la tarea de filtrar información con vistas a aportar calidad al trabajo periodístico. A continuación, se encuentra la contextualización y la combinación,

es decir, el cruce con otras fuentes de información y bases de datos (incluyendo aquí la geolocalización), y finalmente, la difusión del producto, con un fuerte componente visual.

Esta modalidad periodística se encuentra vinculada a otros periodismos, como el de investigación o el de precisión, así como a la documentación, la infografía, el "fact checking", el *open data* y el *big data* (Peiró y Guallar, 2013). Se trata de una práctica que nace del uso de grandes cantidades de información, procedentes de fuentes internas y externas, que pueden ser usadas y difundidas en diferentes formatos, a través de canales diferentes. Y que está en conexión con un gran número de aplicaciones que ejecutan, parcial o completamente, tareas de naturaleza documental, con base en una cultura de datos abiertos, donde la clave no solo es su visualización sino el acceso real.

En el trabajo de García Jiménez y Catalina García (2016), después de analizar diferentes casos de periodismo de datos en *El País*, *ABC* y *Elconfidencial.es*, se observa el número elevado de fuentes de información empleadas, la gran variedad de formatos de documentos, y el trabajo con la base de datos como principal herramienta documental. Destaca el uso de software destinado a la elaboración de mapas y visualizaciones (D3) y de datos geoespaciales (CartoDB), así como el de hojas de cálculos tales como Excel o Google Sreadheets.

En un contexto dominado por la transformación radical hacia redacciones digitales, y de la presencia de los medios sociales, y con una proyección más Search Engine Optimization (SEO) que informativa, se hace más necesario que nunca profundizar en el estudio y prospectiva de la relación entre los periodistas y los documentalistas, en este caso en el dominio de la Documentación Informativa. Y, claro está, hacen necesario conectar las acciones de marcado semántico y los instrumentos como los tesauros con el comportamiento informacional, aspecto tratado en estudios como el trabajo de Pintado (2013).

En definitiva, lo que aquí se plantea es si estos fenómenos pueden conducir a un cambio en el modelo documental existente y, en consecuencia, en los procedimientos y herramientas de la organización del conocimiento. En el sentido que apunta Dowd (2016), la complejidad para la gestión documental se eleva, puesto que los datos, enlazados, están siendo empleados para nuevos cometidos gracias a su capacidad de unir diferentes dominios y diversas plataformas. Una de las posibles respuestas vendría de la mano de lo que Heravi y McGinnis (2015) denominan *Social Semantic Journalism*. Por medio de técnicas como

el procesamiento del lenguaje natural y los metadatos semánticos, se mejora la detección y acceso a la información de actualidad, con capacidad para verificarla y contrastar las fuentes.

Desde nuestro punto de vista, no está en juego el papel de la organización del conocimiento, sino conocer cuáles son las fórmulas adecuadas para enfrentarse no solo al periodismo en red (Sampedro, 2014) sino también al activismo de datos (Milan y Gutiérrez, 2015). En un contexto en el que se deben tener en consideración los cambios en la configuración y estructura de las noticias y la realidad de nuevos espacios por donde éstas se difunden, como los medios sociales. Lo que puede poner en cuestión el paradigma documental tradicional, basado en la transformación en piezas almacenadas derivadas en las bases de datos documentales, que están gestionadas con herramientas documentales al uso.

### 3. Instrumentos y procedimientos

En la línea de lo que señala Moreira (2012), la clave siempre se situará en qué medida se puede representar el conocimiento, a partir de la reducción semántica que se propone, con vistas a la recuperación de información. También sugiere tener presente el trabajo que se hace en la teoría del conocimiento, y añadir los objetivos pragmáticos propios de nuestro dominio. Los cambios que se están produciendo en la formas de producción y difusión de la información y del conocimiento evidentemente afectan a las técnicas y finalidades propias de la representación del conocimiento, tarea que hasta ahora se ha hecho a partir de clasificaciones, tesauros y en algunos casos folksonomías y ontologías.

#### 3.1. Tipos de vocabularios

La primera piedra de toque se circunscribe al tipo de lenguaje que se aplica, ya sea taxonomía, folksonomía, tesoro o incluso ontología. Desde un punto de vista global, hay autores que defienden, dadas las características del ciberespacio, la conveniencia de las folksonomías frente a otras opciones. Es la línea de trabajo defendida por Yedid (2013), que subraya los aspectos positivos de las folksonomías:

- Las etiquetas extraídas se corresponden, en gran medida, con los términos que aparecen en los vocabularios controlados. Globalmente, son similares en los conceptos implicados en las etiquetas únicas y en el uso de sustantivos, así como en la utilización de ortografía reconocida. En cualquier caso, y desde su punto de vista, las inconsistencias, por ejemplo en el plural y singular, no representarían un porcentaje muy alto.

- Se acercan más al vocabulario propio de los usuarios (en lo que se refiere a los términos y a la especificidad deseada), lo que puede tener su influencia positiva en la elección del léxico y la traslación en la recuperación de información de las necesidades de información de los individuos, ya que se convierte en un proceso más intuitivo. A su vez, se suman más voces y puntos de vista a esta actividad: se podría hablar de una mayor democratización al considerar a las etiquetas seleccionadas como conocimiento compartido, y se sustituiría así un enfoque autoritario y centralizado por una visión colaborativa.
- Requiere menos esfuerzo que un vocabulario controlado. Este valor es recomendable para situaciones en las que se dispone de presupuestos bajos. Las folksonomías se muestran más cercanas a lo amateur, si bien podría suponer una penalización para entornos profesionales como el que nos ocupa. En realidad, se recomiendan más cuando no hay profesionales de la documentación cerca, o bien si el número de documentos analizados es considerable. En este último caso, puede incluso ser adecuada la distribución de las actividades indizadoras al mayor número de personas posibles, también a los autores. Aunque puedan surgir problemas a raíz de la falta de conocimiento al respecto del fondo documental de o de los propios objetivos documentales explícitos e implícitos. En definitiva, un recurso puede ser indizado más de una vez por diversos usuarios, lo que da lugar a una descripción intersubjetiva, agregada y quizá más exhaustiva.

En lo que se refiere a las ontologías, y a pesar de las concomitancias existentes con el resto de lenguajes (García Jiménez, 2004), y desde un plano puramente documental, Codina y Pedraza (2011) ya apuntan la consistencia de los tesauros. Son tecnologías ya contrastadas con una trayectoria consistente en los sistemas de información, que se derivan de modelos de trabajo con una implantación positiva desde hace muchos años (Mendes et al., 2015). Esto no quita para que las ontologías, que están detrás de algunos avances en documentación, sean instrumentos consolidados en otras esferas, pero pendientes de desarrollos en nuestro ámbito.

En este punto conviene destacar el uso e implementación de SKOS y su conexión con lenguajes pertenecientes a la web semántica; ya que hace posible que el tesoro se presente en un nivel alto de formalización, lo que además de mejorarlo favorece la interconexión con otros lenguajes y la eficiencia de los sistemas de información. En realidad, en el entorno semántico es cada vez

es mayor la presencia de ontologías como Semantically Interlinked Online Communities (SIOC) o Friend Of a Friend (FOAF). La primera hace posible el uso de datos en línea generados por usuarios y permite nuevas aplicaciones semánticas. Mientras, FOAF permite describir a las personas y las conexiones que tienen entre ellas.

Por su parte, partiendo de una posición relativa a las clasificaciones, Szostak (2014) entiende que las ontologías deben hacer frente a dos problemas serios si quieren una plena integración en la web semántica. La gran cantidad de ontologías, y los diferentes orígenes y supuestos de partida complican su traducción e interoperabilidad. La conexión entre bases de datos por ordenadores a través de ontologías incompatibles se presenta como una dificultad seria. Otra pega viene de la mano de la terminología, puesto que en términos generales es muy vaga. Una posible solución puede provenir del formato de RDF *triplets*, y mediante el esquema de cosas, relaciones y propiedades como base para dar forma a la estructura de la web semántica.

Finalmente, en el trabajo de Soler y Gil (2010) se analizan las diferencias y concomitancias de los diferentes instrumentos señalados. A partir de categorías válidas para el análisis de nuestro objeto de estudio (objetivo, productor, lenguaje con que se trabaja, estructuras que se aplican, los tipos de relaciones, el coste de elaboración y mantenimiento, la actualización, las aplicaciones, los modelos de presentación, la inclusión de definiciones, las propiedades de los términos, el control de la ambigüedad y de la sinonimia, y los editores que facilitan la construcción) concluyen que los diferentes lenguajes, por nivel de complejidad, y de menor a mayor, se escalarían de la siguiente forma: las folksonomías, las taxonomías, los tesauros y las ontologías.

### 3.2. Nuevos espacios

Sánchez, Colmenero y Moreiro (2012) apuntan, durante los últimos años, la relevancia del tesauro y su formalización (lo que condiciona la estructura relacional) en aras de potenciar los vínculos e interoperabilidad de los sistemas de organización del conocimiento (KOS, Knowledge Organization System).

Los instrumentos de organización del conocimiento también son deudores de SKOS en la medida que hacen más fácil su uso y aumentan su eficiencia en conexión con la web semántica. En este sentido, Pastor et al. (2012) analizan el comportamiento de diferentes lenguajes controlados que aplican el modelo de datos Simple Knowledge Organization System (SKOS) en el marco de la web semántica y el *linked open data*. En

cuanto al acceso a los datos, tanto los tesauros como las clasificaciones consiguen buenos resultados en comparación con los encabezamientos de materia y los ficheros de control, que muestran un peor comportamiento. Pero si nos fijamos en el grado de interoperabilidad, los resultados obtenidos por los encabezamientos de materia son mejores, por lo que cabría esperar una mayor integración de SKOS en las bibliotecas. Por último, parece sorprendente el grado bajo de interconexión entre tesauros, lo que quizá se podría solucionar a través de un mayor esfuerzo en las propiedades de mapeado entre conceptos.

Conviene no dejar al margen la potencial relevancia de los gestores de tesauro. En el trabajo de Martínez y Alvite (2014) se aplica un método para evaluar las herramientas de gestión de tesauros a partir de criterios como el objetivo de la aplicación, sus funcionalidades, la capacidad para mantener la integridad y la coherencia del vocabulario, y los aspectos relacionados con la interoperabilidad, especialmente en lo que se refiere a la compatibilidad con estándares propios de la web semántica (RDF/SKOS), la integración con otros sistemas de información, o la capacidad de combinación con otros tesauros. En definitiva, se percibe una clara diferenciación entre las aplicaciones de gestión de tesauros y las aplicaciones orientadas a la web semántica, fundamentalmente a SKOS. Para estos autores, son dignos de mención PoolParty o SKOSed, dada su capacidad para adoptar formatos de web semántica, o bien Turtle y N3, a la hora de importar y exportar de tesauros.

Otro de los aspectos que puede ser de interés es el de la automatización, en diferentes niveles, en la construcción de los tesauros. En el trabajo de Gil y Sánchez (2014) emplean las técnicas de recuperación de información para la generación de tesauros, gracias a las técnicas de cluster para los términos empleados y la categorización, y a las técnicas de similitud para la propuesta de las relaciones semánticas.

### 3.3. Análisis, etiquetado y formalización

En lo que se refiere al análisis y marcado semántico, Pastor (2013) defiende planteamientos que superen el actual papel de Schema.org, siempre teniendo en consideración que el fin último es el de incrementar la efectividad de la recuperación de información. Por ejemplo, el uso de datos estructurados en código X(html) en los sitios web presentan consecuencias positivas tanto en la interoperabilidad semántica, como en la recuperación y la reutilización de la información. De todas formas, lo recomendable sería mirar hacia el modelo que se considere más apropiado tanto para

las tareas analíticas como para su reconocimiento por parte de los motores de búsqueda, ya sea como microdato, microformato o RDFa, o sus combinaciones. Y en cada opción primará la naturaleza semántica de los contenidos, el objetivo de uso, así como la posible reutilización que se plantee.

Según este mismo autor, y conforme al paradigma actual, el posicionamiento de los contenidos está íntimamente ligado al uso de microdatos y al Schema.org. Excepto en aquellos casos que requieren una descripción del contenido muy detallada, en las publicaciones ligadas al *linked open data* y en generación automática de datos a partir de contenidos web, lo que requiere de RDFa. Ante este panorama, posiblemente una solución puede venir de la mano de auténticas redes de equivalencias entre los diferentes modelos de formalización. Son muchos los problemas que se deben solucionar, entre los que destaca la conexión entre el mercado y los vocabularios o la relación con los modelos de consulta y búsqueda.

Por otra parte, y más allá de los modelos de formalización, también destacan fenómenos como el etiquetado social que, a juicio de Pérez et al. (2013), hace progresar la gestión de la información. Desde su punto de vista, este procedimiento permite plantear términos y etiquetas más concretas y con mayor relación con los intereses de los usuarios. Se ajustan más a los entornos compartidos y colaborativos, entendidos como espacios donde se distribuyen libremente etiquetas y recursos. Una de las derivas que tiene este enfoque es la creación de auténticas comunidades o grupos de usuarios, y la combinación con otras formas de comunicación, e incluso de organización social, lo que se puede producir a través de la valoración, positiva o negativa, ante una información o dato.

En cualquier caso, es ya tradición la escasa atención que se presta desde la academia a la selección de términos que, en su momento, servirán de base a los vocabularios y a la recuperación de información. Hecho nunca exento de matices ideológicos y de impacto en la sociedad y los individuos, especialmente en campos como el periodístico.

#### 4. Organización del conocimiento en el periodismo

Las últimas experiencias que se observan en la literatura sobre el particular están referidas, en primer lugar, a nuevos lenguajes documentales. Rubio Lacoba (2012) aborda el caso de *El País*, que goza de una folksonomía controlada, basada en una especie de indexación social que se

asigna a los diferentes productos periodísticos del diario. Sobresale que los periodistas también propongan las etiquetas, con posteriores filtros realizados por los documentalistas al objeto de eliminar tanto la sinonimia como la polisemia y permitir el enriquecimiento con relaciones entre los términos-etiquetas.

La base de trabajo se encuentra en las experiencias del *The New York Times*, *Delicious* y *Soitu*. Se trata de un lenguaje que también se ha nutrido de las base de datos propias, el IPTC en su versión digital, así como de otros tesauros especializados. De igual modo, aplica técnicas SEO para la delimitación de los términos. Con una estructura organizada en las áreas de temas, personajes, organizaciones, lugares y eventos, presentaba 74.000 términos (incluyendo el nombre de personajes) en 2012. El etiquetado se realiza antes de la publicación de la noticias y, por medio de portadillas dinámicas, se vincula de forma automatizada cada noticia a los temas con los pueden plantear algún tipo de conexión. El editor también permite: incluir y vincular nuevas etiquetas; eliminar la ambigüedad fundamentalmente de siglas, cargos y fechas; proponer palabras en comparación con noticias similares; recibir solicitudes de términos por parte de los periodistas; y establecer etiquetas que no se pueden quedar en el lenguaje principal pero que se corresponden con “conceptos editoriales”.

En esta misma línea de modificación de los tesauros, destaca la aproximación de Baños (2013), centrada en la capacidad que ofrecen las noticias de divulgación científica para la renovación y actualización de los tesauros. Por medio de técnicas propias de la recuperación de información (N-gramas y Apache Solr), se plantean equivalencias con diferentes macrotesauros. La primera conclusión de utilidad es que estos macrotesauros no son adecuados para el análisis de noticias orientadas a la divulgación científica, fundamentalmente por la ausencia de términos que resultan. No obstante, queda demostrado que se pueden utilizar como base para la elaboración de otros instrumentos de organización del conocimiento. Las noticias de divulgación resultan apropiadas para dicha actualización, y se constituyen como un auténtico yacimiento de conceptos y términos.

Centrados en el campo de los metadatos y el mercado semántico, Baños et al. (2015) revisan, en el dominio periodístico, los diversos modos de representación de los metadatos, caracterizados por falta de uniformidad, lo que va en contra realmente de los objetivos de estandarización e interoperabilidad. En su estudio se observa que frente a la mayor presencia, en la bibliografía so-

bre el particular, de NewsML y NIFT, lo más habitual es encontrarse con schema.org así como dos esquemas orientados a los medios sociales, que son Twitter Cards y Open Graph Protocol (en Facebook) en los códigos fuente. En menor medida, se acude a normas como rNews, que está preparada para aplicar marcas semánticas a noticias en sistemas de información web, o bien al HTML, el Dublin Core o los microdatos.

Por otra parte, y más allá de estas consideraciones técnicas, García Gutiérrez (2011), que trabaja alrededor de la idea de exomemoria, sostiene la necesidad de aportar una perspectiva pluralista y democrática en las memorias digitales, mediante instrumentos como los tesauros, las ontologías, y las prácticas del análisis documental. Desde un enfoque transcultural y participativo, incide en la relevancia social de las actividades que giran alrededor de las etiquetas, los metadatos, las categorías, las relaciones semánticas. Su mirada se inclina por lo transnacional y transcultural al objeto de superar las limitaciones y enclaustraciones propias de imposiciones positivistas o estereotipadas.

Partiendo de estos postulados, y en el marco del trabajo con mecanismos lógico semánticos y operadores lógico-críticos, propone un análisis documental-semántico en el campo periodístico, capaz también de registrar la información observable y de detectar omisiones, errores y solapamientos (García Gutiérrez, 2013, 2014; García Gutiérrez y Martínez Ávila, 2014)). Su registro modelo se compone de, en primer lugar, una descripción externa, en la que se incluye información sobre el autor, el título, el medio de publicación, la fecha, y la página (o URL). A continuación, plantea la macroproposición, que es una síntesis textual de la acción, considerada la clave de toda la estructura analítica, y planteada mediante un verbo o nombre de proceso. También se incluye el sujeto, que abarca el ser vivo, colectivo o cosa que lleva a cabo la acción que predomina, previamente seleccionada. Y finalmente, el predicado, que se constituye conforme a los aspectos que se deriva de la ejecución de la acción principal definida. Los puntos más relevantes en los que se debe fijar el analista a lo largo del texto son: sujetos, objetos, elementos asociativos, situación, causa, finalidad, consecuencia, modo, instrumento, lugar y tiempo.

## 5. Puntos de análisis

A partir del estado de la cuestión que se ha planteado hasta el momento, proponemos los puntos sobre los que puede realizarse un análisis sobre la organización y representación del conocimiento en la documentación periodística en su apartado

textual. En un primer momento, se debe analizar el contexto teniendo en consideración algunas de las siguientes cuestiones:

- Redacción y convergencia digital: esto tiene que ver tanto con las prácticas periodísticas en sí como con las prácticas digitales cotidianas.
- Características del medio: desde diferentes enfoques que van desde la orientación editorial, el trabajo diferenciado conforme a las plataformas, la presencia en papel, el número de visitas, o las rutinas informativas existentes, entre otros.
- Perfiles profesionales: características de periodistas, programadores, especialistas en medios sociales. Cada vez tienen más influencia las personas encargadas de tareas como la generación de flujo de información o en la aplicación métricas web y su conexión con la documentación y la organización del conocimiento.
- Orientación documental: es relevante detenerse en la existencia de documentalistas y de un servicio de documentación. Interesa detectar funciones y características.
- Usuarios del sistema: es decir, hábitos informacionales globales y necesidades de información.
- Bases de datos: sistema de gestión empleado y diseño de la base de datos.
- Fondo documental: características y naturaleza de la colección de documentos y datos existentes.

En cuanto al vocabulario, los elementos que se tendrán en consideración son los siguientes:

- Tipología: la primera piedra de toque se circunscribe al tipo de lenguaje que se aplica, ya sea taxonomía, folksonomía, tesauro o incluso ontología. O bien combinación de clases.
- Antecedentes y fuentes: se ha de precisar cuáles son los antecedentes, si los tiene, el lenguaje y con qué bases se ha construido.
- Términos: en lo que se refiere a cuestiones como el número, la tipología, el procedimiento de selección, las formas de refinamiento, el nivel de especificidad o el impacto en la recuperación de información.
- Estructura: formas de y existencia de relaciones semánticas y de formas de categorización. Organización conceptual del propio instrumento.

- Productores y usuarios: qué profesionales y en qué forma son los encargados de la elaboración y mantenimiento, y cuáles de su acceso y consulta.
- Actualización: se debe analizar los métodos y periodos de actualización.
- Modelos de interoperabilidad y compatibilidad semántica: conexión en diferentes niveles y sentidos con las normas o propuestas de formalización. También se incluye la conexión con otros vocabularios o instrumentos.
- Gestor: software empleado para la elaboración y mantenimiento del constructo.
- Evaluación y calidad, norma: nivel de seguimiento de las normas vigentes. Seguimiento de la calidad.
- Coste: es importante determinar el coste tanto de la elaboración como del mantenimiento.
- Aplicaciones periodísticas: se trata la virtualidad directa e indirecta en el medio de comunicación.

Por otra parte, y en lo que se refiere al marcado semántico, se debe analizar los siguientes aspectos:

- Elementos y categorías de descripción semántica: cuáles son las categorías que se emplean.
- Tipología de la descripción: con vistas a determinar si se plantea un etiquetado social o un etiquetado profesional.
- Procedimiento de elaboración: análisis, lo más detallado posible, de los mecanismos que se emplean para determinar que un término o etiqueta es válido para la representación.
- Lenguaje: análisis de la relación entre el análisis o marcado y vocabulario correspondiente.
- Recuperación de información: estudio sobre las implicaciones de esta tarea para la recuperación de información y la resolución de las necesidades de información.
- Formalización y codificación: estudio sobre el modelo de formalización que se plantea para esta actividad.
- Calidad y evaluación: hemos de fijarnos en los procedimientos, si existen, para evaluar el resultado del etiquetado o marcado semántico.

## 6. Conclusiones

Tal y como se señalaba al principio de este trabajo, se ha establecido el contexto y la situación de las técnicas e instrumentos que afectan a la organización y representación del conocimiento, en especial a lo que tiene que ver como los tesauros, las folksonomías y los modelos de análisis y marcado semántico. A continuación, se ha analizado la situación de este ámbito en lo que se refiere a la Documentación periodística. Y esto se ha llevado a cabo para delimitar, finalmente, aquellos temas que necesitan de una mayor concreción para dibujar el escenario real de la organización del conocimiento en el cambiante ámbito periodístico.

Como aspectos destacables podemos señalar, por un lado, la continuidad de la creciente tecnificación y de las propuestas vinculadas a la web semántica. No obstante, parece que este proceso choca con la falta de aplicaciones normativas unívocas. En segundo lugar, se debe señalar también el impacto real de lo que podríamos denominar como organización del conocimiento social. No solo se observa en la generalización del uso de los medios sociales, con las consecuencias correspondientes para los sistemas de información, sino también en los lenguajes y el marcado semántico.

## Referencias

- Baños Moreno, María José (2013). Fuentes para la actualización de macrotesauros: noticias de divulgación científica. // Cuadernos de Gestión de Información. 3 (2013) 13-24.
- Baños Moreno, María José; Felipe, Eduardo R.; Pastor Sánchez, Juan Antonio; Martínez Béjar, Rodrigo y Lima, Gercina (2015). Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos. // XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de noviembre, 2015.
- Codina, Lluís; Pedraza Jiménez, Rafael (2011). Tesauros y ontologías en sistemas de información documental. // El profesional de la información. 20:5 (sept-oct, 2011) 555-563.
- Dowd, Cate (2016) The new order of news and social media enterprises: visualisations, linked data, and new methods and practices in journalism. // Communication Research and Practice. 2:1 (2016) 97-110. DOI:10.1080/22041451.2016.1155339
- Ferreras Rodríguez, Eva M. (2012). Nuevos perfiles profesionales: el Periodista de Datos. // Actas IV Congreso Internacional Latina de Comunicación Social. [http://www.vistalatinacs.org/12SLCS/2012\\_actas/062\\_Ferreras.pdf](http://www.vistalatinacs.org/12SLCS/2012_actas/062_Ferreras.pdf)
- García Gutiérrez, Antonio (2011). Epistemología de la Documentación. Barcelona: Stonberg, 2011.
- García Gutiérrez, Antonio (2013). Entrevistas periodísticas y textos declarativos: un procedimiento de análisis y recuperación documentales basado en modos de búsqueda pronominal. // El profesional de la información. 22:4 (julio 2013), 315-325

- García Gutiérrez, Antonio (2014). Análisis documental de noticias de prensa en sistemas de información factual. // Revista Española de Documentación Científica. 37:2 (abril-junio, 2014) e046.
- García Gutiérrez, Antonio; Martínez Ávila, Daniel (2014). Critical organization of knowledge in mass media information systems. // Knowledge organization. 41:3 (2014) 205-216
- García Jiménez, Antonio (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // Anales de Documentación. 7 (2004), 79-95. <http://revistas.um.es/analesdoc/article/view/1691/1741>
- García Jiménez, Antonio y Catalina García, Beatriz (2016). Procesos y herramientas documentales para el periodismo de datos. // 1º Congreso Internacional sobre Periodismo de Datos. Nuevas narrativas para el periodismo especializado. Málaga, 25-26 de febrero de 2016 (en prensa).
- Gil Urdiciain, Blanca; Sánchez Jiménez, Rodrigo (2014). Técnicas de recuperación de información aplicadas a la construcción de tesauros. // TransInformação. 26:1 (jan-abr, 2014) 19-26.
- Heravi, Bahareh Rahmanzadeh; McGinnis, Jarred (2015). Introducing Social Semantic Journalism. // The Journal of Media Innovations. 2.1 (2015) 131-140
- Martínez González, M. Mercedes; Alvite Díez, M. Luisa (2014). Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. // Anales de Documentación. 17:1 (2014).
- Milan, Stefania y Gutiérrez, Miren (2015). Medios ciudadanos y big data: la emergencia del activismo de datos. // Mediaciones. 14 (enero-junio, 2015) 10-26.
- Mendes, Paula Raphisa; Martins dos Reis, Raquel y Coura Moreira dos Santos Maculan, Benildes (2015). Tesauros no acesso à informação: uma retrospectiva. // Revista ACB: Biblioteconomia em Santa Catarina. 20:1 (2015) 49-66.
- Moreira, Walter (2012). Relações conceituais como ponto de inflexão entre as linguagens documentais, a terminologia e as ontologías. // Scire, 18:2 (2012), 123-127.
- Pastor Sánchez, Juan Antonio (2013). Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. // Scire. 19:2 (2013) 55-68.
- Pastor Sánchez, Juan Antonio; Martínez Méndez, Francisco Javier; Rodríguez Muñoz, José Vicente (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. // El profesional de la información. 21:3 (may-jun, 2012) 245-253.
- Peiró, K. y Guallar, J. (2013). Introducció al periodisme de dades. Característiques, panoràmica i exemples. // Item. 57 (2013) 23-36
- Pérez Sanchidrián, Elaine; Campos Posada, Raúl, y Campos Posada, Gloria Elisa (2014). Etiquetado social: un modelo de representación de la información en la blogosfera. // Biblos. 56 (2014) 19-27.
- Pintado Navarro, Rocio (2013). El comportamiento informacional de los periodistas en la Región de Murcia. Cuadernos de Gestión de Información. 3 (2013) 25-51.
- Rubio Lacoba, María (2012). Nuevas destrezas documentales para periodistas: el vocabulario colaborativo del diario El País. // Trípodos. 31 (2012) 65-78.
- Saad Corrêa, Elizabeth y Bertocchi, Daniela (2012). A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. // MATRIZES. 5:2 (ene-jun, 2012) 123-144
- Sampedro, Víctor (2014). El cuarto poder en red. Por un periodismo (de código) libre. Barcelona: Icaria, 2014.
- Sánchez Cuadrado, Sonia; Colmenero Ruiz, María Jesús y Moreiro, José Antonio (2012). Tesauros: estándares y recomendaciones. // El profesional de la información. 21:3 (may-jun, 2012) 229-235.
- Soler Monreal, Concha y Gil Leiva, Isidoro (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. // Revista Interamericana de Bibliotecología. 33:2 (jul-dic, 2010) 361-377.
- Stone, Martha L. (2014). Big data for Media. Reuters Institute for the Study of Journalism. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media\\_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf) (23-02-2016)
- Sugimoto, Cassidy R.; Ding, Ying and Thelwall, Mike (2012). Library and Information Science in the Big Data Era: Funding, Projects, and Future [a panel proposal]. // Proceedings of the ASIST 2012. 49:1 (2012), 26-31.
- Szostak, Rick (2014). Advances in Classification Research Online 2013. Classification, Ontology, and the Semantic Web. // Advances In Classification Research Online. 24:1 (2014) 30-37.
- Yedid, Nadina (2013). Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. // Información, Cultura y Sociedad. 29 (dic. 2013) 13-2.

---

Enviado: 2016-03-31. Segunda versión: 2016-07-18.  
Aceptado: 2016-07-18.

---