



Trabajo Fin de Máster

Creación y diseño de un tesoro propio para
informativos de televisión

Creation and design of thesaurus for television news

Autoras

M^a Inmaculada Hervás Moreno

Ana C. Lafuente Gómez

Directora

Piedad Garrido Picazo



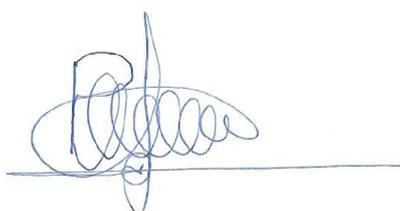
Facultad de
Filosofía y Letras
Universidad Zaragoza

Creación y diseño de un tesoro propio para informativos de televisión

M^a Inmaculada Hervás Moreno

Ana C. Lafuente Gómez

V^o B^o de la directora



Fdo: Piedad Garrido Picazo

Firma de las autoras



Fdo: M^a Inmaculada Hervás Moreno

Ana C. Lafuente Gómez

HERVÁS MORENO, M^a Inmaculada

Creación y diseño de un tesoro propio para informativos de televisión / M^a Inmaculada Hervás Moreno, Ana C. Lafuente Gómez; directora Piedad Garrido Picazo – Zaragoza : Universidad de Zaragoza, Facultad de Filosofía y Letras, 2017. – 66 p.: 29 cm. – Trabajo de Fin de Máster de Consultoría de Información y Comunicación Digital 2017.

1. Ciencias de la Información – TFM. 2. Documentación – TFM. I- Lafuente Gómez, Ana C. II- Garrido Picazo, Piedad (dir.). III- Creación y diseño de un tesoro propio para informativos de televisión.

02: 025.4

CREACIÓN Y DISEÑO DE UN TESAURO PROPIO PARA INFORMATIVOS DE TELEVISIÓN

M^a Inmaculada Hervás Moreno

Ana C. Lafuente Gómez



Imagen 1. Helen Louise Brown.

*Primera persona en emplear el término tesauro
en el campo de la Información y Documentación en 1957.*

Resumen

El presente Trabajo Final de Máster (TFM) surge de la necesidad de ETIQMEDIA, empresa especializada en la catalogación supervisada de contenidos audiovisuales, de crear una herramienta para controlar el lenguaje utilizado en los informativos de televisión, en otras palabras, facilitar el trabajo de minutado y análisis de los informativos, para: (i) mejorar la posterior Recuperación de Información (RI) del contenido textual, (ii) posibilitar la realización de un correcto análisis de dicho contenido y, (iii) obtener estadísticas de uso de términos concretos en tiempo real.

Tras estudiar otras alternativas existentes en el mercado como pueden ser las listas de términos, las taxonomías y las ontologías, se vio claro que lo que necesitaba la empresa era un tesoro propio, ya que es la solución que proporciona más posibilidades de cara a la (i) indización, (ii) navegación, (iii) futura RI y, (iv) el trabajo de la no obvia e importante tarea del establecimiento de relaciones de sinonimia, así como la clasificación de los términos por secciones, que permitan clasificar cada minuto de un informativo.

Para finalizar, comentar que en su construcción han sido fundamentales herramientas y arquitecturas tales como Excel, TemaTres, SKOS y ETIQMEDIA-Indexing & Broadcast, aplicación propia de la empresa, donde se ha integrado nuestra propuesta de tesoro, denominado EM-Informativos/ETIQMEDIA, que con 3.308 términos en su haber, cubre por el momento todas las necesidades informativas de la empresa, estando en consonancia con otros tesoros similares consultados.

Palabras clave

Informativos, metadatos, SKOS, TemaTres, tesoro.

Abstract

The current Final Master`s Degree Project arises from the need for ETIQMEDIA, a company specialized in the supervised cataloging of audiovisual content, to create a tool to control the language used in television news, in other words, facilitate the work of minuted and information analysis, in order to: (i) improve the subsequent Information Retrieval (IR) of the textual content, (ii) enable a correct analysis of this content and (iii) obtain use statistics of specific terms in real time.

After studying other alternatives in the market such as lists of terms, taxonomies and ontologies, it was clear that what the company needed was a thesaurus of its own, since it is the solution that provides more possibilities for the (i) indexing, (ii) navigation, (iii) future IR, and (iv) work on the non-obvious and important task of establishing synonymy relationships as well as classifying terms by sections, that allow classifying every minute of an informative.

To conclude, comment that in its construction have been fundamental tools and architectures such as Excel, TemaTres, SKOS and ETIQMEDIA-Indexing & Broadcast, own application of the company, where has been integrated our proposed thesaurus, called EM-Informativos / ETIQMEDIA, which with 3,308 terms under its credit, covers for the moment all the information needs of the company, being in line with other similar thesauri consulted.

Keywords

Metadata, news, SKOS, TemaTres, thesaurus.

Índice

1	<i>Introducción</i>	7
1.1	Justificación	7
1.2	La Documentación y el Periodismo	9
1.3	Historia de los Tesoros	12
1.4	Definición de Tesoro	15
1.5	Estado de la cuestión	24
1.6	Características comunes de los tesauros	30
2	<i>Tesoro de Informativos</i>	32
2.1	Objetivos	34
2.2	Herramienta	35
2.3	Organización	40
2.4	Creación	42
2.5	Aplicación	55
3	<i>Conclusiones</i>	59
4	<i>Licencia documental</i>	62
5	<i>Bibliografía</i>	63

1 Introducción

1.1 Justificación

La realización de este tesoro propio para informativos de televisión surge de la necesidad de la empresa aragonesa ETIQMEDIA¹, de crear una herramienta para controlar el lenguaje que se utiliza en los informativos de las principales cadenas de televisión en España (TVE, Antena 3, Cuatro, Telecinco, La Sexta y 13TV). Así, el objetivo del mismo era la facilitación del trabajo de minutado y análisis de informativos, siendo especialmente importante el establecimiento de las relaciones de sinonimia y la clasificación de los términos por secciones que permitieran posteriormente clasificar cada minuto de un informativo televisado.

Con toda esta información acumulada, documentada y automatizada la empresa podría realizar tanto el análisis sobre el contenido como extraer estadísticas sobre el uso de términos concretos o temáticas abordadas, disponiendo así de una valiosa e importante información tanto para Agencias de Comunicación y/o Publicidad como Consultorías Audiovisuales. A esta relación cabría incluir también como posibles interesados a las Instituciones públicas y empresas privadas que desearan conocer al detalle el impacto de su marca y su gestión en la ciudadanía.

¹ ETIQMEDIA, empresa de creación y desarrollo de tecnología al servicio del contenido audiovisual especialmente en materia de catalogación y subtitulado. www.etiqmedia.com

A este respecto, cabe destacar, tal y como señalan numerosos estudios sobre los usos y consumos de los medios de comunicación², que la televisión sigue siendo el medio de comunicación de masas dominante, si bien, como matiza Castells (2009) ahora las tecnologías, los negocios y la cultura la han transformado profundamente, hasta el punto de que ahora se considera un medio que combina difusión masiva con difusión personalizada, lo cual hace más importante y necesario conocer de primera mano los usos y consumos de la misma para ir adaptando los contenidos a la demanda de los usuarios.

En lo concerniente al origen de la creación de este tesoro, la empresa ETIQMEDIA lanzó una oferta de contratación dirigida a estudiantes en prácticas, de la que resultaron seleccionadas las alumnas Hervás Moreno y Lafuente Gómez.

Así, tras analizar las características de creación del tesoro, las autoras de este Trabajo Fin de Máster (en adelante TFM), junto con el profesor Miguel Ángel Esteban Navarro y la profesora Piedad Garrido Picazo (apoyadas y animadas también por otras profesoras) decidimos asumir el reto de realizar el TFM de forma conjunta acogiéndonos al art.1 Definición, Modalidades y Características del Reglamento de Fin de Máster³. De esta forma, y siguiendo con la normativa del Reglamento citado, se inició la tarea de convertir las prácticas del *Máster Universitario en Consultoría de Información y Documentación Digital* (en adelante MUCID) en el cuerpo de este TFM, adoptando una de las modalidades que permitía el Máster a la hora de realizar la memoria o proyecto: “Trabajos específicos realizados como resultado de prácticas en empresas o instituciones” (art.1.4).

A favor de las autoras se contaba principalmente con el tándem humano y profesional formado por una documentalista y una periodista, además de ser conscientes de que se trataba de la realización de un trabajo pionero y único hasta hoy, reto que supuso un estímulo desde el minuto cero. Asimismo, la realización de este proyecto nos iba a

² Se realiza una búsqueda en Google con los términos: “Estudios sobre el uso y consumo de los medios de comunicación”. Se obtiene el siguiente resultado: “Aproximadamente 12.200.000 resultados (0,62 segundos). [Consulta realizada 07/09/2017].

³ Reglamento de los Trabajos Fin de Grado y Fin de Máster de las titulaciones que se imparten en la Facultad de Filosofía y Letras. Aprobado por acuerdo de la Junta de la Facultad de 4 de marzo de 2015.

permitir poner en práctica una buena parte de los conocimientos adquiridos durante la realización del máster y culminar el proceso de aprendizaje llevado a cabo a lo largo de la titulación. Además, cabe añadir que dentro del MUCID cada una de las alumnas cursó una especialidad diferente del Máster: Hervás Moreno realizó la “Especialidad en Comunicación Corporativa” y Lafuente Gómez la “Especialidad en Servicios de Información”.

Esperamos que a través de la realización de este TFM hayamos conseguido mostrar y demostrar de forma integrada y compacta, algunos de los contenidos formativos recibidos y las competencias y habilidades adquiridas en relación con el contexto en el que las hemos aplicado.

1.2 La Documentación y el Periodismo

En pleno siglo XXI nadie pone en cuestión que la información es poder. Dos palabras, información y poder, que desde siempre han viajado juntas en el tiempo cual binomio inseparable. Un binomio al que en este TFM se ha añadido conscientemente un tercer vocablo por la importancia que tiene en el proceso informativo como es la documentación, ya que coincidimos con Caldera (2015) en señalar que todos los medios de comunicación diseñan e implementan los sistemas de gestión documental pensando en los periodistas.

Por lo que se refiere a la información, y reafirmando el poder que tiene, cabe señalar que este empoderado término en el momento actual está experimentando una seria crisis de credibilidad debido principalmente al extendido síndrome de la infoxicación ciudadana, infoxicación que, según lo define la Fundación del Español Urgente (en adelante FUNDEU⁴), consiste en la incapacidad de análisis eficiente de un flujo de información elevado.

La ingente cantidad de información a la que tenemos acceso las personas es de tal magnitud que, el mismo ejercicio de querer estar mínimamente bien informado, conlleva realizar un esfuerzo superior de análisis de lo publicado al objeto de poder

⁴ FUNDEU. Disponible en: <https://tinyurl.com/ycd45d7p>

separar el trigo (la información contrastada) de la paja (información de dudoso origen), esfuerzo intelectual que, por otra parte, poca gente está dispuesto a hacer. Toda esta información que circula libre y principalmente por Internet, crea en los usuarios una cierta “ilusión de sabiduría y conocimiento” que a su vez impide al lector detenerse a reflexionar y asimilar sobre lo que está leyendo debido a la velocidad que se genera la misma. Además, cuanto más rápida fluye la información, más se acelera el proceso de caducidad de ésta, creándose una sensación de constante renovación global de ideas, noticias y opiniones.

Por otra parte, en cuanto al término poder, se ha creído oportuno recoger la definición actualizada de este concepto expresada por el sociólogo Manuel Castells por el certero análisis y concreción del mismo: “el poder es la capacidad relacional que permite a un actor social influir de forma asimétrica en las decisiones de otros actores sociales de modo que se favorezcan la voluntad, los intereses y los valores del actor que tiene el poder” (Castells, 2009:33).

Se coincide con el autor en señalar que, en la sociedad de la información actual, el concepto de actor se refiere a distintos sujetos de la acción, es decir, actores individuales y/o colectivos, organizaciones, instituciones (públicas y privadas) y redes, cuyo fin último no sólo es comunicar lo que hacen sino también posicionarse para tratar de imponerse en la comunicación.

En esta relación de poder-información, Castells llega a afirmar que la influencia más importante en el mundo de hoy es la transformación de la mentalidad de la gente. “Si esto es así –añade-, los medios de comunicación son las redes esenciales, ya que ellos, organizados en oligopolios globales y sus redes de distribución, son la fuente principal de los mensajes y las imágenes que llegan a la mente de las personas” (Castells, 2009:55).

Y, en cuanto al peso e importancia de la documentación y los procesos derivados de ella, en la actual sociedad en red, se puede decir que es el pilar fundamental para poder acceder, disponer e incluso analizar una buena parte de la información de una forma lógica y sencilla.

En otras palabras y siguiendo a Naumis (2007), se puede decir que el éxito en la difícil empresa de prestar un adecuado servicio de información y referencia, estará determinado en buena medida por la organización documental. Además, a este respecto cabe recordar, siguiendo a Agirreazaldegui (1996:87) que la Documentación encuentra sus bases científicas en la Teoría General de la Información y Comunicación: “La Documentation n’est qu’une des branches d’une classe plus générale: les moyens d’Information et de Communication” (Otlet y Wouters 1934)⁵.

Desde entonces hasta nuestros días, numerosas han sido las definiciones que conectan de forma inexorable la información con la documentación. De todas ellas se expone la de García Gutiérrez (1996) al ser ésta una de las primeras en formularse y describir la doble vertiente de la documentación: por un lado, está al servicio de la investigación de las Ciencias de la Información y, por otro lado, se aplica a la realización eficaz y rigurosa del trabajo informativo en los medios de comunicación. “Las Ciencias de la Información conforman la fundamentación de la Documentación desde la perspectiva teórica: la Documentación se declara Ciencia y práctica informativas entroncando, su magnitud teórico-conceptual, en la epistemología de la Información y de la Comunicación. Las leyes y modelos que rigen los procesos comunicativos son válidos en los procesos documentales que se confiesan abiertamente modos y medios de comunicación social” (García G., 1996:22).

Por último, se ha creído conveniente cerrar este apartado sobre la importancia de la relación entre información-poder-documentación recuperando algunas de las conclusiones de la profesora y novelista Chacón Gutiérrez plasmadas en su tesis doctoral -defendida en 1994-, sobre “La fiabilidad del proceso documental del discurso periodístico como fuente de información: análisis, método y valoración”. En ella, la autora parte de la base de que la Documentación es un medio de comunicación masivo que es capaz de persuadir y manipular, si no se realiza con profesionalidad y objetividad.

⁵ Citado por Sagredo e Izquierdo (1983). Traducción: “La Documentación es solo una de las ramas de una clase más general: los medios de Información y de Comunicación”.

En el desarrollo y justificación de la tesis, recupera las afirmaciones de García Gutiérrez y Lucas Fernández (1987) “los archivos de noticias constituyen las memorias vivas de la historia reciente” y de Otlet (1989), “los documentos constituyen en su conjunto la memoria materializada de la humanidad”, señalando que para que esto se produzca “será necesario que el usuario se acerque a los archivos de noticias con la confianza absoluta en que le van a proporcionar documentos relativos a informaciones verídicas”, centrando en ello la investigación.

La tesis de Chacón (1994) concluye que “analizando de forma desglosada la fiabilidad del Proceso Documental del discurso periodístico con respecto al emisor, con respecto al receptor, y con respecto al mensaje, se obtienen suficientes datos para valorar cuantitativamente la certeza de fiabilidad y los grados de fiabilidad óptima, fiabilidad absoluta y fiabilidad relativa, tanto de forma global (fiabilidad del proceso) como individualmente (en referencia a sus tres elementos: tomando como base para el análisis al emisor, el mensaje o al receptor)” (1994:389).

En definitiva, se puede señalar que corresponde al periodista, al profesional de la **información**, no sólo dar a conocer los hechos sino acompañar éstos con razonamientos críticos, contextualización histórica y temporal y datos que inviten a la reflexión, utilizando para ello la **documentación** necesaria, dejando a un lado el sensacionalismo y el maniqueísmo en aras a ofrecer una información rigurosa que permita a la ciudadanía crear sus propias opiniones.

1.3 Historia de los Tesauros

Antes de abordar las diferentes definiciones que se han realizado sobre el término tesoro, se ha considerado oportuno hacer una breve reseña de la historia y procedencia de los tesauros.

A este respecto, siguiendo al grupo de expertos de “Vocabularios estructurados para la recuperación de la información” del Comité 50 de AENOR en la publicación de García Marco (2007) cabe recordar que los primeros tesauros se desarrollaron dentro del ámbito científico-técnico en los años cincuenta y sesenta al objeto de superar las

dificultades que planteaba al proceso de la Recuperación de la Información (en adelante RI), el solo empleo de palabras clave y, posteriormente de descriptores.

Los teóricos de la documentación coinciden en señalar en que el primer tesaurus publicado fue el de “DuPont de Nemours” en USA en 1959, y los primeros en ser utilizados por una amplia base fueron el “Chemical Engineering Thesurus” publicado en 1961 por el American Institute Of Chemical Engineers, y el “Thesaurus of ASTIA Descriptors”, publicado por el US Defense Documentation Center en 1962. La herramienta alcanzó su madurez en 1967 con el “Thesaurus of Engineering and Scientific Terms” del Engineers Joint Council y el US Departament of Defense, expandiéndose rápidamente en el contexto internacional.

Martín Gavilán (2009) fija en este mismo período de tiempo de comienzos de los años 60 la creación de los primeros tesaurus y lo justifica señalando que la explosión de la información, su diversificación y especialización, desbordaron la capacidad de los lenguajes clasificatorios que se venían desarrollando hasta entonces para representar y recuperar el contenido. “La aparición del sistema Unitérmino de Taube (1951-1953), la elaboración del índice KWIC de Luhn (1959) y la indización por conceptos (descriptores) de Mores (1960) revolucionaron las técnicas de indización y pusieron las bases teóricas y metodológicas para la creación en los años 60 de los primeros tesaurus”, (2009:21).

Abundando en los impulsores de los primeros tesaurus, varios teóricos entre los que se encuentra López Alonso (1998) señalan que fue Eugenio Wall quien definió los principales contenidos lingüísticos de los tesaurus documentales en materia de sintaxis, semántica, género, sentido, etc.

Asimismo, la comunidad científica coincide en afirmar que fue la investigadora, informatóloga y documentalista Helen Louise Brownson quien utilizó el término tesaurus por primera vez el 14 de mayo de 1957 en la “Dorking Conference on Classification”, citándolo como una herramienta para la RI. En dicha contribución, el término tesaurus es utilizado al analizar la problemática de traducir los conceptos y sus relaciones, tal y como se expresan en los documentos, a un lenguaje con mayor

precisión y sin ambigüedades, para facilitar la RI (Arano, 2005). Así y según Brownson, los tesauros se incorporan en los sistemas de recuperación de información desde los años cincuenta con la idea de “transformar los conceptos y sus relaciones que se expresan en los documentos, en un lenguaje regularizado, con los sinónimos controlados y las estructuras morfosintácticas simplificadas” (Brownson, 1957)⁶.

En España, la primera vez que se empleó el término tesoro fue a cargo del jurista y obispo de Huesca don Vidal de Canellas en su libro “*Compilatio Maior o In excelsis Dei Thesaurus*”, obra en la que se compila y comenta el derecho aragonés, y que fue realizada con motivo de la celebración de la Corte General en Huesca en el año 1247.

Si nos ceñimos a la época contemporánea del siglo XX, para el caso español y siguiendo a Currás (2005 y 2012) cabe señalar que en nuestro país “en el ámbito referido a los estudios teóricos sobre tesauros, hemos de reconocer que hasta la década de los años 80 del siglo XX no se ha notado una gran actividad” (2005:89). Para esta investigadora, los casos aislados como el de Javier Lasso de la Vega en su *Manual de Documentación*, o la celebración de REUNIBER-78⁷ –cuya labor repercutió en los años sucesivos-, o las propias contribuciones de Currás a Congresos o Simposios, son la excepción de la regla.

A partir de este momento, existe un amplio consenso en destacar la importancia de la creación de las cátedras de Lingüística Documental y de Teledocumentación en las Facultades de Ciencias de la Información y de las Facultades y Escuelas Universitarias de Biblioteconomía y Documentación para potenciar el uso, creación y estudio de los tesauros.

Para cerrar este apartado sobre la historia de los tesauros, cabe señalar que simultáneamente, y unida a esta evolución, hay que subrayar la importancia que ha tenido -y sigue teniendo en materia de creación de tesauros-, el impacto tecnológico digital desde la aparición de Internet y de sus aplicaciones asociadas, convirtiéndose éstas en una prolongación de la actividad cotidiana del ser humano.

⁶ Este texto es una de las primeras referencias al uso de los tesauros como herramientas de apoyo a la recuperación de información. Tomado de Fernández-Pamillón, 2010.

⁷ REUNIBER-78. Conferencia Iberoamericana sobre Información y Documentación Científica y Tecnológica celebrada en Madrid en 1978.

A este respecto, bien se podría afirmar que la World Wide Web (en adelante WWW) es la responsable de la aparición de nuevas formas de relacionarse surgidas en la sociedad, especialmente en todo el ámbito de la comunicación, desde el proceso de emisión de información hasta el de recuperación y acceso de la misma.

Cada vez el ser humano quiere saber más, encontrar información más concisa y específica, y conseguirla en el menor tiempo posible. Y en este campo, la tecnología es la mejor aliada. En palabras de Arano (2005): los cambios tecnológicos están conduciendo a una progresiva digitalización del ámbito de la representación y recuperación de información que afecta por igual a los recursos de información, las herramientas de representación y recuperación, y los requerimientos de los usuarios.

1.4 Definición de Tesoro

Para desarrollar este apartado, dado que las autoras de este TFM no son expertas en la materia, en consenso con la directora del mismo, se ha decido presentar una relación sobre la evolución y diferentes definiciones del término tesoro que ya han sido consensuadas y/o aceptadas por la comunidad científica, teniendo en cuenta que la mayor parte de los teóricos de la Documentación y de la Información han realizado su propia definición.

Para ello, y tras la lectura y consulta de abundante bibliografía se citarán las de algunos investigadores e investigadoras como muestra representativa ya que en todas ellas existen matices distintos que se complementan. Entre estos teóricos se encuentran⁸ a Agustín Lacruz, Alvite Díez, Arano Poggi, Caro Castro, Cordón García, Currás Puente, Fernández Quijada, Gil Urdiciain, Gilchrist, López-Huertas, Martín Gavilán, Martínez Méndez, Martínez Usero, Monchón Bezares, Naumis Peña, Pastor-Sánchez, Rodríguez Bravo, Rodríguez López, San Segundo o Sorli Rojo, entre otros.

Asimismo, cabe añadir que también los diferentes estándares de normalización ofrecen su propia definición, convirtiendo en misión (casi) imposible el objetivo de reflejar y condensar todas las definiciones en un solo documento, teniendo en cuenta

⁸ Nota autoras. Esta relación se presenta en orden alfabético como único criterio de ordenación.

además, que esa exhaustividad excedería ampliamente los propósitos del presente TFM.

A este respecto, y al objeto de ilustrar lo comentado, así como la ingente bibliografía que hay diseminada por la red sobre este término, se aporta la captura de pantalla realizada en Google el pasado 7 de septiembre. Como se puede observar, en poco menos de un minuto, el buscador universal arrojó una cifra de 407.000 resultados de la búsqueda “definición de tesoro” (Imagen 2).

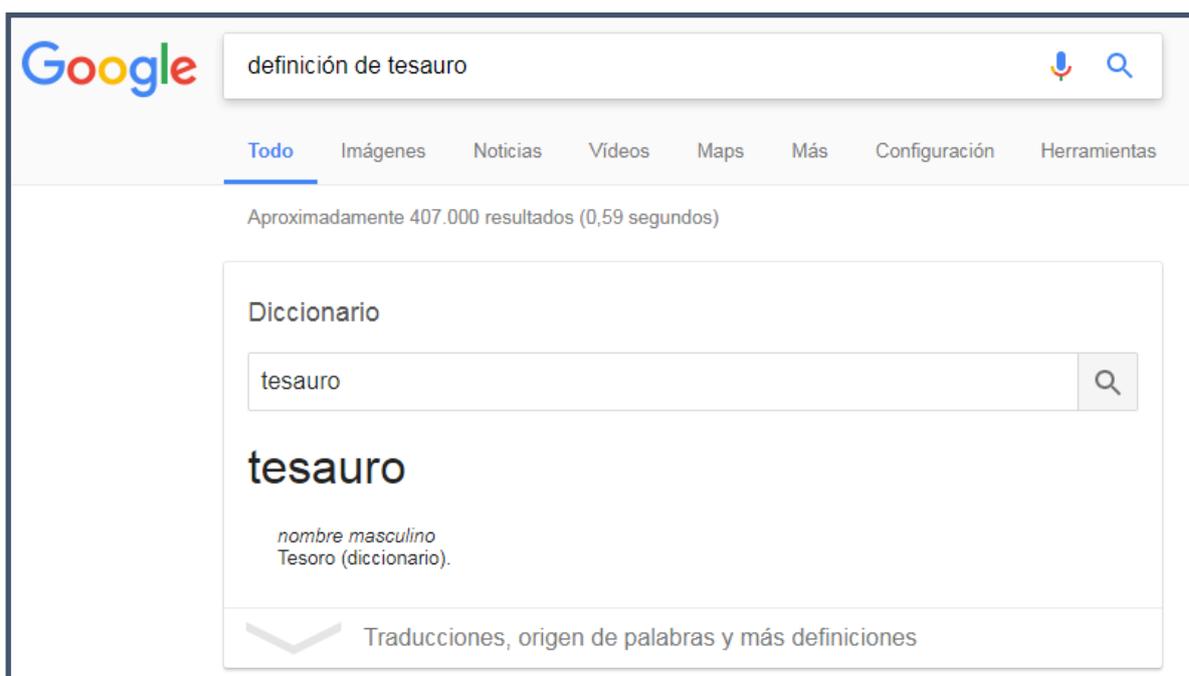


Imagen 2. Captura pantalla búsqueda término tesoro. (07/09/2017)

Como señalan Naumis (2007), Currás (2005, 2012), y todos los especialistas en esta materia, la palabra tesoro, en su origen thesaurus o thesauros, era el nombre dado en las casas de época griega a la cámara donde se guardaba el tesoro. Posteriormente se usó para designar ciertos diccionarios, índices o repertorios de términos, con la misma connotación de lugar donde se guardaba lo más importante, pero con el matiz de que en este caso y debido a su evolución, lo que se guardaba era parte del saber y del lenguaje humano (Martínez de Sousa, 1993).

Las definiciones encontradas, como se puede observar, incluyen desde la que recoge la Wikipedia hasta las propuestas por las Normas internacionales estandarizadas, pasando por las de algunos teóricos de la Información, en general, y de la Documentación en particular.

A continuación se presenta una relación de las definiciones encontradas y que hacen referencia a la segunda mitad del siglo XX y principios del siglo XXI. A excepción de la definición encontrada en la Wikipedia, el resto supusieron en su época un punto de inflexión y de avance en la definición del término, sirviendo de base para la mayoría de los estudios posteriores.

En la presente relación de definiciones ordenada cronológicamente se puede observar cómo ha ido evolucionando el término tesouro a medida que se investigaba y perfeccionaba su creación y aplicación.

Se comienza la exposición, con la definición que aporta la Wikipedia ya que se ha considerado que es la principal fuente consultada, al realizar por primera vez una aproximación al término tesouro. Además, hay que señalar que es la mejor posicionada en Internet.

“Wikipedia⁹. En Biblioteconomía y Ciencias de la Información, Tesouro es un sistema de organización del conocimiento, en forma de lista, que contiene los «términos» controlados empleados para representar los conceptos, temas o contenidos de los documentos, con miras a efectuar una normalización terminológica que permita mejorar el canal de acceso y comunicación entre los usuarios y las Unidades de Información (entiéndase unidad de información como: biblioteca, archivo o centro de documentación). El Tesouro es un intermediario entre el lenguaje que encontramos en los documentos (lenguaje natural) y el que emplean los especialistas de un determinado campo del saber (lenguaje controlado). Aunque los incluye, las entradas de un tesouro no deben ser consideradas solo como una lista de sinónimos”.

⁹ Localizado en <https://es.wikipedia.org/wiki/Tesouro> [Consulta 09/09/2017]

La siguiente definición está firmada por Emilia Currás y data de los años 70 en sus inicios como profesora en la Universidad Autónoma de Madrid. Esta doctora, documentalista e informatóloga solía aportar a su alumnado una definición de tesoro que, a su parecer, era “muy intuitiva y muy adecuada para iniciar el tema” (Currás, 2005:81). Para ella:

“Un tesoro es un vocabulario especializado en el que las palabras que lo componen están relacionadas, unas con otras, semánticamente y sintácticamente”.

En esta época, en 1971, aparecen dos obras importantes que fueron referente indiscutible en los siguientes avances del desarrollo de los tesauros: una en el Reino Unido de Alan Gilchrist y otra en la antigua República Federal de Alemania de Gernot Wersig. A este respecto, se presenta en primer lugar la de Wersig (1971) y más adelante se hará lo propio con la de Gilchrist, puesto que en el año 2000 este teórico presentó una definición más completa del término.

“Wersig (1971). Este informático y documentalista alemán define al tesoro como una lista de términos, prefijados con anterioridad pero sacados del texto de los documentos, que desdobla los conceptos en las unidades simples. Estas se coordinan posteriormente para evitar ambigüedades. Entre ellas se establecen relaciones jerárquicas, asociativas y de equivalencia”¹⁰.

Siguiendo este orden cronológico, incluimos ahora la definición aportada por la UNESCO:

“UNESCO (1976). En 1976 se publicaron los manuales de la UNESCO dentro del programa UNISIST, donde se definen los tesauros según su función (instrumento de control terminológico, usado para trasladar, desde un lenguaje natural de los documentos, los descriptores, a un sistema lingüístico) y según su estructura (son vocabularios controlados y dinámicos de términos

¹⁰ Citada en Soler (2009:12).

relacionados semántica y genéricamente, que cubren un dominio específico del conocimiento)".

En ese mismo año, F. Levéry (1976)¹¹, formuló una de las definiciones más concisas realizadas hasta entonces:

"Levéry señaló que el tesoro es un puente entre el lenguaje del informado (el documentalista) y el lenguaje del no informado (el usuario)".

A continuación añadimos la definición de Long (1980), que también supuso un importante impulso en la concreción y descripción del término. Para él:

"Un tesoro es un conjunto semiológico, utilizado como sistema de clasificación, donde las unidades clasificatorias forman conjuntos de morfemas, ligados por relaciones de tipo paradigmático, a partir de las cuales se construyen diferentes clases. Estos morfemas hacen uso de relaciones de tipo sintáctico para lograr que los términos tengan una representación documentaria. Lo importante no son los documentos, son las relaciones de los temas. Esta tarea la realiza el indizador" (obtenido de Currás, 2005:82).

Se cierran las aportaciones de los teóricos del siglo XX con la de Martínez de Sousa, quien introduce el término "informática" en la definición del mismo:

"Martínez de Sousa (1995): Un tesoro es un vocabulario limitado, generalmente de palabras especializadas, dotado de sus correspondencias semánticas, y elegido para que represente las nociones que figuran en un texto dado para su empleo en informática y en el establecimiento de índices".

En esta relación, se comienza el siglo XXI con la aportación anunciada en párrafos anteriores del teórico inglés Gilchrist:

"Gilchrist (2000), en la 4ª edición de su obra sobre construcción de tesauros, publicada en el año 2000 escribe que tesoro es un vocabulario de un lenguaje de indización controlada, de forma que "a priori" se hacen explícitas las

¹¹ Extraído de Chaumier (1986).

relaciones entre conceptos, para ser usadas en los sistemas de recuperación de información, saltando desde el catálogo de fichas a Internet”.

La teórica Blanca Gil Urdiciáin incluye en su definición algunas de las ventajas que contienen los tesauros:

“Gil Urdiciáin (2004) describe los tesauros de la siguiente forma: un tesoro es un lenguaje documental de estructura combinatoria, de carácter especializado, que se basa en expresiones conceptuales llamadas descriptores, provistas de sus relaciones semánticas. Y cita como ventajas del mismo su flexibilidad, su capacidad de especialización, la multiplicidad de combinaciones que permite establecer entre los términos de su vocabulario, el alto nivel de control terminológico y su facilidad de revisión y puesta al día”.

De ese mismo año, se aporta la definición de los especialistas Aitchinson y Clarke quienes inciden en la cualidad de los tesauros de “ayudar a las personas o las máquinas” a encontrar la información que buscan:

“Aitchinson y Clarke (2004): Los tesauros son herramientas lingüísticas que sirven para ayudar a las personas o a las máquinas a encontrar los términos más apropiados para expresar una idea. Sus aplicaciones más frecuentes son de tipo lingüístico, documentalista, informático y académico”.

En el primer lustro del siglo XXI, aparecen las definiciones que presentan la Norma estandarizada ANSI/NISO y la investigadora Silvia Arano, ambas de 2005.

“ANSI/NISO Z39.19 (2005), define el tesoro como un vocabulario controlado organizado y estructurado de forma conocida y donde las relaciones entre términos de equivalencia, homógrafos, jerárquicas y asociativas se visualizan claramente mediante marcadores estándares y recíprocos”.

“Arano (2005). Para esta técnica e investigadora, un tesoro es un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento. Proporciona una organización semántica principalmente a través de la explicitación de las relaciones

establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los términos que los representan. Asimismo, los tesauros son herramientas construidas para el auxilio tanto de los profesionales de la información como de los usuarios finales”.

Asimismo, se ha creído interesante aportar la definición de tesouro de la bibliotecóloga Naumis Peña:

“Naumis Peña (2007). Se puede definir el término tesouro en la disciplina de bibliotecología como sistema de clasificación temática o facetada, cuya estructura básica está conformada por una relación de descriptores que representan o describen autoridades o contenidos temáticos, mediante unidades lingüísticas, semánticas y sus relaciones, extraídas del lenguaje formal de una disciplina o área específica del conocimiento (2007:128)”.

En la recta final de esta exposición sobre la evolución de la definición del término tesouro se citan dos aportaciones aparecidas el mismo año, 2009, y la definición de la Norma UNE-ISO (2011).

En la definición realizada por César Martín Gavilán se incluye una de las características más relevantes de los tesauros según su criterio: la representación unívoca de los conceptos.

Martín Gavilán (2009:22): “El tesouro es un prototipo de lenguaje de indización y recuperación controlado, que se basa en la postcoordinación de sus descriptores. Además representa de manera unívoca los conceptos de los documentos evitando así los problemas de homonimia, sinonimia, polisemia (relaciones de significado), así como el establecimiento de las relaciones jerárquicas y de relación entre los descriptores que lo componen”.

Como fruto de su tesis doctoral sobre la evaluación de los vocabularios controlados, la doctora Soler Monreal (2009) aporta no sólo la definición de tesouro sino que incluye también uno de sus principales objetivos: satisfacer las necesidades de información.

“Soler Monreal (2009) define a los tesauros como vocabularios controlados, de estructura combinatoria, definidos a priori, es decir, fijados con anterioridad, compuestos por términos que reflejan conceptos que se relacionan entre sí semántica, jerárquica y asociativamente, que tienen como finalidad el control de sinónimos, y que se utilizan para describir de manera unívoca, traduciendo del lenguaje natural al documental, el contenido de los documentos, para su posterior recuperación en un sistema documental dado, con el fin de satisfacer las necesidades de información”.

La penúltima definición que se presenta es la de la Norma UNE-ISO ya que desde la aparición de los primeros tesauros hasta hoy, la normativa que regula su elaboración, objetivos, usos y gestión ha ido evolucionando y adaptándose sobre todo teniendo en cuenta el entorno digital. De las distintas normas existentes como la ISO 2788 de 1986 y la UNE 50-106 de 1990, reflejamos en este TFM la correspondiente a 2011 por ser la última en incluir definición de tesoro, y por lo tanto la más actualizada:

“Norma UNE-ISO 25964-1 (2011) explica que un tesoro se trata de un conjunto de conceptos seleccionados que están representados principalmente por términos preferentes organizados formalmente para que las relaciones paradigmáticas entre los conceptos se hagan explícitas, y los términos preferentes se acompañen de sinónimos o cuasi sinónimos (2011:10)”.

Por último, se cierra esta muestra de la evolución del término tesoro con la definición incluida en el recién creado Tesoro de Mujeres (2014) realizado por la Biblioteca de Mujeres de Madrid tras cuatro años de trabajo, a cargo de Mediavilla y Folla (2014).

En la introducción del mismo, queda reflejada no solo la definición del término tesoro sino también la importancia de estas herramientas de documentación e información. Así, las autoras señalan que: “un tesoro es algo más que un instrumento para organizar y buscar información. Es un conjunto de términos que refleja y representa el mundo en que vivimos, una sociedad que, bajo la pretendida universalidad y neutralidad del conocimiento masculino, ha negado y niega la participación de las mujeres en la historia y en la construcción social”, (2014:3).

En esta misma línea se habían pronunciado con anterioridad Caro y San Segundo (1999), si bien estas investigadoras se refieren a los lenguajes documentales. Para ellas, los lenguajes documentales deberían facilitar el acceso a la información a todos los usuarios estableciendo una estructura lógica y coherente que mejore su interacción con el sistema. Sin embargo el resultado obtenido en algunas de sus investigaciones “evidencia que los criterios ideológicos prevalecen sobre los lógicos y que, además, estos criterios suponen una discriminación para determinados grupos sociales”, (1999:101).

Para las investigadoras, esta cuestión no baladí, está mejor resuelta con los tesauros. Añaden y refuerzan su planteamiento señalando que las relaciones asociativas entre términos están presentes de manera desigual en los diferentes lenguajes documentales ya que “en las clasificaciones enumerativas tradicionales predominan las relaciones sintagmáticas (jerárquicas, verticales) sin que aparezcan apenas términos asociados o equivalentes, y los tesauros son los que, al menos en teoría, presentan una estructura semántica plenamente desarrollada”, (1999:102).

Por último, se retoma la definición de tesoro incluida en el Tesoro de Mujeres, dando por finalizado este apartado. Para las autoras del mismo:

“Martínez y Folla (2014) un tesoro es un vocabulario estructurado de un determinado campo del conocimiento. Es un lenguaje natural controlado para organizar, almacenar y recuperar la información; instrumento utilizado por documentalistas para analizarla e indizarla y por usuarias y usuarios para buscarla, las dos funciones básicas del tesoro. Está integrado por descriptores y no descriptores. Un descriptor es un término del lenguaje natural que expresa, de forma breve, el tema del que trata el documento”, (2014:32).



Imagen 3. Tesauro de Mujeres (disponible en: <http://tinyurl.com/yaoz825c>)

1.5 Estado de la cuestión

Aun teniendo en cuenta la importancia que tienen en nuestra sociedad los medios de comunicación en general y la televisión en particular, tras acometer este TFM resulta sorprendente constatar que el número de trabajos similares al presente en cuanto a profundización en la estructura, fuentes y herramientas de documentación, no ha sido muy abundante. En la siguiente imagen, se observan los resultados de la búsqueda realizada en Google (12/08/2017).



Imagen 4. Resultados búsqueda tesauros de medios de comunicación

Entre algunas razones expuestas, Caldera Serrano (2015) señala que “en cierta forma tiene su lógica derivada de que los grandes avances se realizaban en el marco de las empresas audiovisuales, que en ningún momento querían compartir ni estandarizar su gestión documental”, situación similar a la del presente tesoro puesto que se trata del encargo de una empresa de consultoría mediática.

Entre las investigaciones realizadas más completas se encuentra la de Castillo y De la Cueva (2007) de hace una década, y en ella se sustenta principalmente el presente apartado. En la investigación se hace un recorrido nacional e internacional de la situación y uso de los lenguajes controlados en documentación informativa y sobre todo de tesauros especializados en las tres modalidades informativas: prensa (edición impresa y digital), radio y televisión.

Como principales conclusiones señalan que: la documentación periodística adolece de instrumentos adecuados de clasificación e indización de la información, a excepción

del *Subject Reference System* del *International Press Telecommunications Council* (en adelante IPTC)¹²; que cada medio de comunicación utiliza su propio lenguaje controlado, elaborado independientemente, y “en muchas ocasiones se trata de vocabularios de bajo nivel, como simples listas”; y que, los estudios sobre el empleo de lenguajes controlados en medios de comunicación españoles evidencian un uso predominante de glosarios y sistemas clasificatorios, además de una escasez en la elaboración y uso de tesauros debidamente estructurados (Castillo; de la Cueva, 2007).

A nivel internacional, se afirma que el tesoro de prensa más destacado por todos los autores es el empleado en la base de datos del diario *The New York Times*, que data de 1971 y puede considerarse como el primero de los de información de actualidad. Incorpora más de 700.000 descriptores e incluye una relación alfabética de nombres de personas.

En cuanto a los medios audiovisuales en España, Castillo y de la Cueva (2007), recogen diversas investigaciones realizadas sobre los tesauros de las cadenas de televisión de Telecinco, RTVE, Televisió de Catalunya (TV3) y Euskal Telebista (ETB).

Los investigadores Martín-Muñoz y López-Pavillard (1995) citan algunos de los tesauros, temáticamente independientes (geográfico, de deportes, de agricultura, de animales, de ciencia y técnica), que se emplean en las bases de datos de documentación audiovisual de Radiotelevisión Española (RTVE) y que fueron creados y desarrollados por el propio centro de documentación a lo largo de su actividad.

Respecto al tesoro de Telecinco, Del Valle, García y un equipo amplio de colaboradores en primer lugar, y posteriormente Del Valle y García (2003), se encargan de realizar el estudio, la construcción y puesta en funcionamiento del mismo en su interesante estudio de caso. El trabajo realizado permite que los usuarios puedan explorar entre la información contenido en los más de cuarenta campos de la base de datos y especialmente entre los 13.5000 documentos (entre programas, brutos y fichas principales) analizados.

¹² En 1999 el IPTC desarrolló la *Subject Reference System* para permitir a los proveedores de información (principalmente agencias) acceder a un lenguaje independiente codificado para describir el contenido de las noticias y para facilitar su intercambio. En Castillo y de la Cueva. (2007).

Según los autores, el tesouro para materiales audiovisuales de Telecinco se construye en respuesta a una iniciativa de los Centros de Documentación de esta cadena de televisión ante la necesidad de integrar su documentación “en un entorno digital muy innovador y ante la constatación de la necesidad de disponer de una herramienta de lenguaje ágil y abierta capaz de tratar todo el conjunto de documentos generados por un sistema de producción de contenidos dirigido a la información y el entretenimiento”, (Del Valle; García, 2003:166).

Para la construcción del tesouro de Telecinco, los temas provisionales fueron extraídos de los códigos propuestos por el IPTC para la transferencia internacional de noticias. Las áreas temáticas que incluye son las siguientes: Arte. Cultura. Espectáculos / Justicia / Desastres y accidentes / Economía. Negocios. Finanzas / Educación / Medio ambiente / Salud / Trabajo / Ocio. Estilo de vida / Política / Religión. Creencias / Ciencia y Tecnología / Asuntos Sociales / Deporte / Guerras. Conflictos.

En la televisión catalana, TV3, Llobet y Pañella (1988) describieron el proceso de elaboración del tesouro para la base de datos de imágenes. Está estructurado en dos partes: la primera incluye nombres propios de personas y entidades (diccionario de autoridades) con relaciones de sinonimia y notas explicativas. La segunda está dedicada a nombres comunes y geográficos que incluyen relaciones jerárquicas (genéricos y específicos), asociativas (términos relacionados) y de sinonimia. Este mismo modelo de tesouro se emplea también para la descripción de imágenes en la *Unidad de Documentación de Radiotelevisión Valenciana*.

Por lo que respecta al *Centro de Documentación y Archivo* de ETB, Castillo; de la Cueva (2007) señalan que para la recuperación temática de su base de datos de imágenes, utilizan un tesouro elaborado por su propio equipo de documentalistas a partir del de TV3. “Consta de 12.000 términos en total: 3.500 son descriptores temáticos, 2.500 geográficos y 6.000 identificadores o descriptores onomásticos” (2007:621).

El hecho de no tener un tesouro propio la mayoría de los medios de comunicación, no significa que no utilicen ninguno. En estos casos, lo que sí que es frecuente es que algunos medios de comunicación en España adapten y usen tesouros ya elaborados

como el de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (en adelante UNESCO) y/o EuroVoc, principalmente.

Por lo que se refiere al primero de los citados, cabe señalar que se trata de un macrotesauro elaborado por la UNESCO¹³ de carácter multidisciplinario disponible en cuatro idiomas: inglés, francés, español y ruso. La versión en español contiene poco más de 8.600 términos y está en constante actualización. Está dividido en ocho temas principales (educación, información y comunicación, ciencias sociales, cultura y humanidades, ciencia y tecnología, política, derecho y economía), y 88 microtesauros. La consulta se puede realizar seleccionando un tema y un microtesauro o buscando distintas cadenas de caracteres en la lista permutada o en la lista alfabética. La información sobre cada descriptor incluye enlaces a los términos que conforman sus relaciones semánticas, el microtesauro al que pertenece y su traducción a las otras tres lenguas. La edición más reciente (sexta ed., 2a rev) es de 2007.

The screenshot shows the SKOS (Tesauro de la UNESCO) interface. At the top, there's a navigation bar with the SKOS logo and the title 'Tesauro de la UNESCO'. Below this, there are language selection buttons for 'Español', 'English', 'Français', and 'Русский', along with a search box labeled 'Texto a buscar'. A sidebar on the left contains a menu with options like 'Introducción', 'Consultar el tesauro', 'Presentación alfabética', 'Presentación jerárquica', 'Punto de acceso SPARQL', 'Descargas', 'Estadísticas', 'Créditos y aviso legal', and 'Vocabulario UNESKOS'. The main content area displays the entry for 'Medios de comunicación de masas' with its URL, a search link, and lists of related terms in different languages and related concepts.

SKOS
Tesauro de la UNESCO

Introducción
Consultar el tesauro
Presentación alfabética
Presentación jerárquica
Punto de acceso SPARQL
Descargas
Estadísticas
Créditos y aviso legal
Vocabulario UNESKOS

Español English Français Русский

Texto a buscar

RDF/XML | N-Triples | N3/Turtle | JSON | JSON-LD

Medios de comunicación de masas
<http://skos.um.es/unescothes/C02425>
Buscar en UNESDOC

Otros idiomas

- Mass media (English term)
- Médias de communication (Terme français)
- Средства массовой информации (Русский термин)

Área → Microtesauro

- 5 Información y comunicación > 5.20 Industria de la información

Usado por (no descriptores)

- UP Medios de comunicación

Conceptos específicos

- TE Medios sociales

Conceptos relacionados

- TR Cine
- TR Comunicación de masas
- TR Cultura de masas
- TR Educación sobre medios de comunicación
- TR Medios de información
- TR Prensa
- TR Radio
- TR Televisión

Imagen 5. Tesauro UNESCO (Medios de comunicación de masas)

¹³ Disponible en <http://databases.unesco.org/thessp/>

El EuroVoc¹⁴ es el tesouro empleado para la confección de los índices del Diario Oficial de las Comunidades Europeas. Se caracteriza por ser un tesouro multilingüe y multidisciplinario que abarca la terminología de los ámbitos de actividad de la Unión Europea (en adelante UE). Contiene términos en 23 lenguas oficiales de la Unión Europea, así como en tres lenguas de países candidatos a la adhesión a la UE. Es administrado por la Oficina de Publicaciones, quien aplica una gestión de tesouros basada en ontologías y tecnologías de web semántica acordes con las recomendaciones del Consorcio World Wide Web (en adelante W3C) y las últimas tendencias en materia de normalización de tesouros. Utilizan el tesouro EuroVoc, entre otros, el Parlamento Europeo, la Oficina de Publicaciones, parlamentos nacionales y regionales de toda Europa, administraciones nacionales, medios de comunicación y usuarios privados tanto de los países miembros de la UE como de terceros países.



Imagen 6. Tesouro EuroVoc

¹⁴ Disponible en: <http://eurovoc.europa.eu/drupal/?q=es/node>

Algunos investigadores, entre los que se encuentran Perpinyà (1995) y García Gutiérrez (1999) que han propuesto la realización de un tesoro conjunto y único para los medios de comunicación han llegado a la conclusión de que es un proyecto poco viable. Entre las razones que esgrimen señalan que la construcción de un tesoro de información de actualidad (el vocabulario periodístico queda obsoleto rápidamente) requiere una alta exigencia de dedicación y tiempo para su elaboración y mantenimiento, con el consiguiente incremento de coste económico que no todos los medios de comunicación están dispuestos a pagar o pueden permitirse.

En definitiva, y como conclusión de este apartado, se puede señalar que las herramientas de documentación en los medios de comunicación audiovisuales son, en pleno siglo XXI, todavía escasas: “actualmente los medios de comunicación en España no cuentan con la utilización de tesauros adecuados y propios” (Castillo y de la Cueva, 2007).

De ahí que, la realización de este tesoro de informativos de televisión pueda llegar a significar un punto de inflexión en el avance y extensión de este tipo de herramienta fundamental para los medios de comunicación.

1.6 Características comunes de los tesauros

A lo largo de la historia de los tesauros, si bien cada uno de ellos fue creado para responder a objetivos concretos que cubrieran las necesidades para las que fueron concebidos, ello no es óbice para poder señalar una serie de características comunes. A este respecto, y siguiendo a varios investigadores, entre los que se encuentran López Yepes (2000), Naumis (2007) o Currás (2012), podemos señalar algunas características comunes a todos:

- El tesoro es un sistema no independiente porque interactúa en un sistema de información
- Es un producto humano porque está diseñado por el hombre para describir y explicar los contenidos temáticos documentales
- Los elementos que lo componen son constantes y variables

- Cada tesouro tiene un contexto particular que es necesario estudiar para que responda a las necesidades a las cuales está destinado
- Cada tesouro tiene sus propios objetivos, orden sistemático, índices y términos que lo conforman
- El cometido final es recuperar la información asentada en ellos a través de diferentes puntos de acceso
- Los tesouros necesitan una constante revisión y actualización

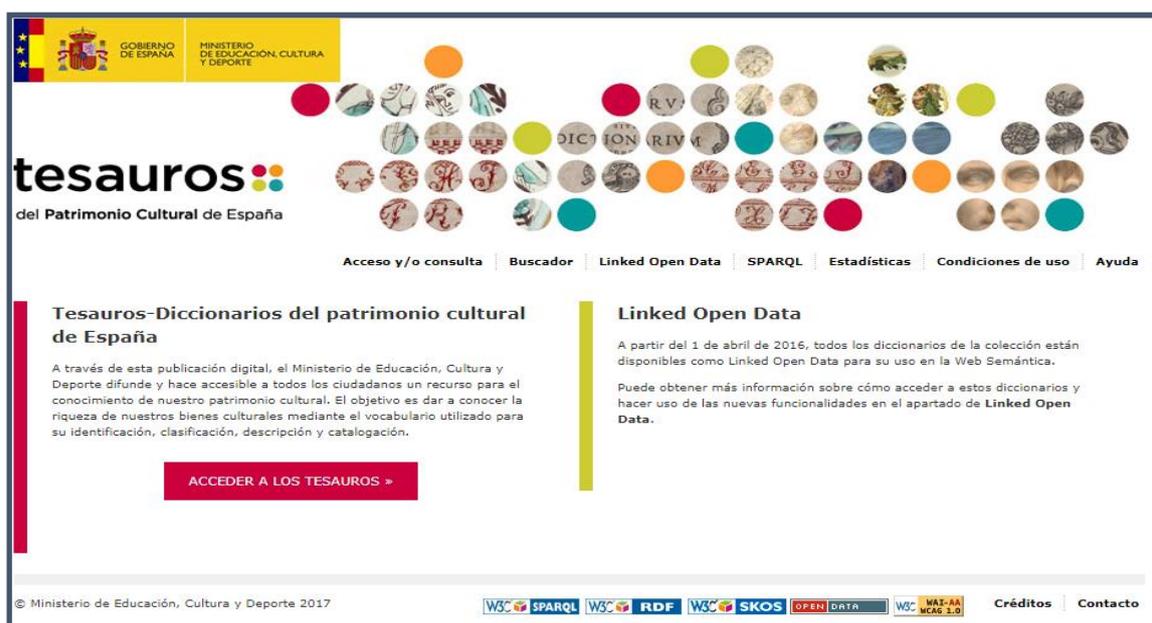


Imagen 7. Tesauros de España. Fuente: www.tesauros.mecd.es/tesauros

2 Tesoro de Informativos

El encargo para realizar este tesoro parte de la necesidad de la empresa ETIQMEDIA de crear una herramienta para controlar el lenguaje que se utiliza en los programas informativos de televisión de las cadenas españolas. Concretamente, en la clasificación del contenido textual de estos para su posterior recuperación. De esta forma, se posibilita la realización de análisis sobre el contenido y estadísticas de uso de términos concretos.



Imagen 8. Logotipo empresa ETIQMEDIA (fuente: www.etiqmedia.com)

La fuente principal para construir este tesoro de informativos de televisión ha sido el conjunto de términos de indización extraídos de los propios textos de los informativos emitidos.

Como fase previa de documentación, durante un año (03/2016-03/2017) se grabaron todos los informativos de las principales cadenas de televisión en España: Televisión Española (TVE), Antena 3, Cuatro, Telecinco y La Sexta. Una vez obtenida esa información, mediante un programa de reconocimiento de voz, se configuró una base de datos de alrededor de un millón de términos. El siguiente paso fue la creación del tesoro en base a estas palabras.

La documentación para este tesoro se circunscribe al campo de los programas de informativos, y concretamente a los boletines diarios de las diferentes cadenas que se emiten en dos franjas horarias determinadas y fijas cada día -a mediodía y por la tarde- y que coinciden con los *prime time* de las cadenas.

Estos programas tienen unas características propias que se diferencian del resto de formatos de otros programas dedicados a la información como pueden ser *Los Desayunos* de TVE, *Espejo Público* de Antena 3, o *Al rojo vivo* de La Sexta. Entre las características comunes de los informativos cabe señalar:

- Franja horaria de emisión
- Estructura de contenido
- Temática de contenido
- Duración
- Equipo humano de realización

HORARIOS EMISIÓN INFORMATIVOS					
CADENAS ----- HORARIOS					
Mañanas	Telediario 1	Antena 3 Noticias 1	Noticias Cuatro	Informativos Telecinco	La Sexta Noticias
	15:00h a 16:00h	15:00h a 15:45h	14:15h a 14:45	15:00h a 16:00h	14:00h a 14:55h
Tardes	Telediario 2	Antena 3 Noticias 2	Noticias Cuatro	Informativos Telecinco	La Sexta Noticias
	15:00h a 16:00h	15:00h a 16:00h	15:00h a 16:00h	15:00h a 16:00h	15:00h a 16:00h

Tabla 1. Franjas horarias de emisión de los informativos (elaboración propia)

Por lo que respecta a la estructura de los informativos cabe señalar que las cinco cadenas de televisión mantienen formatos similares. En cuanto a contenidos, los informativos están divididos en tres bloques: Informativos, El Tiempo y Deportes. En la franja horaria matinal, TVE, Antena 3, Telecinco y La Sexta desarrollan el informativo con el siguiente esquema: Informativos → Deportes → El Tiempo. En el caso de Cuatro el esquema es: Informativos → El Tiempo → Deportes.

Sin embargo, en el informativo que se emite por la noche, coinciden todas las cadenas en hacer dos grandes bloques: Informativos y el Tiempo, incluyendo la sección deportes en el primero de ellos.

2.1 Objetivos

El objetivo de este Tesauro, por tanto, es la facilitación del trabajo de minutado y análisis de informativos, siendo especialmente importante el establecimiento de las relaciones de sinonimia y la clasificación de los términos por secciones que permitan posteriormente clasificar cada minuto de un informativo.

Se pueden enumerar las funciones del tesauro de la siguiente manera:

- Reunir bajo un solo descriptor todas las posibles formas de nombrar a una misma entidad
- Facilitar la búsqueda en los contenidos clasificados por medio de los descriptores y los términos asociados a ellos
- Servir como herramienta para analizar los contenidos de las noticias y realizar análisis y estadísticas

El ejemplo del Presidente del Gobierno de España es el que mejor visualiza la problemática, ya que en el texto de la noticia se puede nombrar al personaje de varias maneras diferentes por: (i) nombre completo, (ii) apellido, (iii) nombre de su cargo, e incluso (iv) si la noticia hace referencia a su partido, por el cargo que ostente en éste. En la figura 1, se presentan los términos asociados al descriptor *Rajoy Brey, Mariano*. Los que tienen la etiqueta UP (Utilizado Por) son los términos asociados que se contabilizarán bajo el descriptor, los de la etiqueta (TR Término Relacionado), son los términos con los que se ha decidido relacionar el descriptor y TG (Término Genérico) son los que clasifican el término en categorías.

<p>Rajoy Brey, Mariano</p> <p>TR: Presidente de Gobierno #2011-</p> <p>TR: Presidente del Partido Popular</p> <p>TR: Presidente del Partido Popular #2004-</p> <p>TR: Presidente de Gobierno</p> <p>TG: Entidad</p> <p>TG: Política nacional</p> <p>TG: Ámbito nacional</p> <p>UP: Mariano Rajoy</p> <p>UP: Mariano Rajoy Brey</p> <p>UP: Rajoy</p> <p>UP: Rajoy Brey</p>
--

Figura 1. Presentación del termino Rajoy Brey, Mariano

2.2 Herramienta

La elección de la herramienta para la edición del tesaurus se realizó tras llevar a cabo un arduo análisis de las opciones disponibles que reunían las condiciones que se habían establecido de antemano.

La empresa requería una comparación y posterior selección entre Protegé, TemaTres y Swoop, además de un análisis de la arquitectura SKOS¹⁵ (Simple Knowledge Organization System) y las posibilidades que proporciona ésta para la posterior interacción de los sistemas automatizados.

En una reunión al inicio de este proyecto (febrero de 2017), se desestimaron tanto la primera (Protegé) como la última (Swoop) puesto que son editores más orientados a la

¹⁵ Iniciativa del W3C (World Wide Web Consortium) en forma de aplicación de RDF que proporciona un modelo para representar la estructura básica y el contenido de esquemas conceptuales como listas encabezamientos de materia, taxonomías, esquemas de clasificación, tesaurus y cualquier tipo de vocabulario controlado.

construcción de ontologías que de tesauros, por lo tanto se centró toda la atención en TemaTres.

Tras su análisis se puede comprobar que se trata de una opción válida y adecuada por las siguientes razones:

- Se trata de una herramienta conocida y utilizada con anterioridad, por las alumnas de máster involucradas en este proyecto
- Cumple con las premisas planteadas por la empresa
 - edita tesauros, los muestra y navega por ellos
 - posibilita la exportación a SKOS
 - cumple con las premisas del software libre y/o gratuito

En la evaluación con un Modelo de Calidad¹⁶ de Uso de software *ad hoc*, desde el punto de vista del usuario final, a la aplicación web TemaTres se le analiza la usabilidad, la calidad de contenido, la visibilidad y la eficiencia, como características de primer nivel. La siguiente tabla contiene la definición y explicación de éstas.

En el Modelo de Calidad se establece la siguiente puntuación (tabla 2), que deben cumplir las características analizadas.

El umbral para la métrica es un valor entre 0 y 1, resultado de aplicar la siguiente operación, donde X es igual a la “puntuación atributo evaluada (de modelo de calidad) dividida por la puntuación del atributo en la aplicación web” y los resultados se comparan con:

- a) $(X = 0)$ -> No hay defectos
- b) $(0 < X \leq 0.3)$ -> Defecto leve
- c) $(0,3 < X \leq 0.6)$ -> Defecto medio
- d) $(0.6 < X \leq 1)$ -> Defecto grave

¹⁶ A partir del estándar ISO/IEC 25010 (SQuaRE).

Característica	Definición	Explicación
Usabilidad	Es el término empleado para denotar la facilidad con que las personas pueden utilizar una herramienta (sistemas informáticos, interfaces, programas de escritorio, aplicaciones web, periféricos, etc.).	TemaTres posee dos tipos de interfaz: una para navegar, consultar y visualizar un tesoro (interfaz pública) y otra para crear, gestionar, administrar y exportar un tesoro (interfaz de administración), por lo tanto la usabilidad debe apreciarse en las dos caras, entendiendo si se cumplen los requisitos que pretende la aplicación, su facilidad de uso y si es clara y segura.
Calidad de contenido	La calidad de la información vendrá determinada, por su capacidad para satisfacer las necesidades de información de las personas que lo utilicen y/o consulten.	En esta aplicación las funcionalidades son las necesarias para crear un lenguaje controlado y en este punto se incide sobre si éstas dan el resultado requerido, por ejemplo las relaciones de equivalencia y las etiquetas UF, UP y USE que guían al usuario hacia el término preferente, por lo que se incide en si están claramente descritas y son fácilmente aplicables en la incorporación de términos.
Visibilidad	La facilidad de acceso y reconocimiento de los contenidos, productos y servicios ofertados en Internet, garantizando la usabilidad, accesibilidad, calidad y localización de la información por cualquier usuario.	Es importante conocer datos sobre la visibilidad de las aplicaciones Web, en este modelo se va a hacer hincapié en la popularidad de la aplicación.
Eficiencia	Característica que nos indica la capacidad de lograr los objetivos deseados con exactitud y exhaustividad, utilizando el menor número de recursos posibles	Es imprescindible conocer el tiempo de carga real de una aplicación Web, y ver qué tareas se pueden realizar para mejorarlo.

Tabla 2. Características de Modelo de Calidad de Software para TemaTres

Características / Atributos	Métrica del modelo de calidad	Puntuación de TemaTres
Usabilidad	40	20
Calidad de contenido	30	18
Eficiencia	10	5
Visibilidad	20	10
Total	100	53

Tabla 3. Métrica para el Modelo de Calidad de TemaTres

Por lo tanto el cálculo de los criterios es:

- a) Usabilidad – $20 / 40 = 0.5 \rightarrow$ defecto medio
- b) Calidad de contenido – $18 / 30 = 0.6 \rightarrow$ defecto medio
- c) Eficiencia – $5 / 10 = 0.5 \rightarrow$ defecto medio -
- d) Visibilidad – $10 / 20 = 0.5 \rightarrow$ defecto medio

Dada la validez de TemaTres se exponen a continuación las propiedades óptimas para este proyecto:

- Creación de relaciones entre términos
 - de equivalencia
 - jerárquicas
 - asociativas
- Redacción de notas para cada término
 - de alcance
 - históricas
 - de catalogación
 - bibliográficas
- Flujo de estados para términos candidatos, aceptados y rechazados
- Soporte para vocabularios polijerárquicos y multilingües
- Cantidad ilimitada de niveles jerárquicos y términos no preferidos
- Buscador de términos repetidos y de términos libres
- Administración y gestión de usuarios y editores
- Estadísticas
- Exportación a: Zthes, RDF, SKOS-Core, Topic Maps, MADS, Dublin Core y BS9723
- Visualización de múltiples niveles de términos por pantalla
- Navegación sistemática y alfabética

En cuanto a la gestión del vocabulario controlado, TemaTres define dos tipos de usuarios (**Editor y Administrador**) sobre los que poder realizar, a posteriori, una auditoría de términos a través de un sumario de actividad basado en la fecha de

creación, última modificación y último cambio de estado. Las funciones tanto del usuario con perfil de **Editor** como del usuario con perfil de **Administrador**, son las que se detallan a continuación:

— Usuario **Editor**

- Accede a todas las funciones de gestión de términos
- Crear, modificar, eliminar términos
- Crear, modificar, eliminar relaciones entre términos
- Crear, modificar, eliminar relaciones entre términos y términos de vocabularios de referencia
- Asignar mapeos terminológicos entre términos del vocabulario y términos de vocabularios de referencia a través de servicios Web
- Actualizar mapeos terminológicos entre términos del vocabulario y términos de vocabularios de referencia a través de servicios Web
- Crear, modificar, eliminar notas
- Modificar sus datos de identificación
- Acceder a datos de auditoría propios
- Generar reportes terminológicos simples y avanzados

— Usuario **Administrador**

- Accede a todas las capacidades del editor
- Accede a las funciones administrativas
- Modificar la configuración del vocabulario controlado
- Crear, modificar, eliminar usuarios
- Registrar, modificar, eliminar vocabularios de referencia
- Exportar versiones completas del vocabulario controlado
- Realizar operaciones de mantenimiento de la base de datos
- Acceder a datos de auditoría totales
- Importar vocabularios desde archivos de texto o SKOS
- Registrar vocabularios de referencia a través de servicios Web

La instalación del paquete de software de TemaTres, disponible en sourceforge.net¹⁷, se debe realizar en un servidor propio que soporte:

- Apache, IIS, etc
- PHP (4.3.0 o superior)
- Un Sistema Gestor de Bases de Datos (SGBD). Por defecto, TemaTres utiliza MySQL

Por lo tanto la instalación del software se realiza en el servidor de la empresa ETIQMEDIA y se crean cuentas de usuario **Editor** para las dos alumnas y de **Administrador** para la empresa. El tesoro se denomina *EM-Informativos / ETIQMEDIA*.

2.3 Organización

Con los requerimientos de la empresa y las posibilidades que ofrece la herramienta elegida para su edición, el tesoro se organiza en una clasificación diseñada previamente, en la que se decide que los términos clasificadores sean meta-términos. Estos son conceptos que no deben utilizarse para la indización, ya que son términos que describen a otros. Cabe recordar que los meta-términos funcionan como guías, facetas, categorías, etc. Se trata, pues, de un tesoro polijerárquico, ya que un concepto mantiene relaciones jerárquicas con más de un término o árbol de términos.

El diseño de la clasificación se realiza en cuatro grupos:

- El primero clasifica los términos en la “Temática de informativos”
 - Internacional
 - Política nacional
 - Sucesos
 - Economía
 - Ciencia, tecnología y salud

¹⁷ SourceForge es un sitio web de colaboración para proyectos de software, que actúa como un repositorio de código fuente, hospedado por VA Software Corporation que es la creadora de la Red de Desarrollo de código abierto.

- Sociedad
 - Cultura
 - Deporte
 - Información meteorológica
- El segundo lo hace por el “Ámbito” en el que se puede desarrollar la noticia
- Ámbito internacional
 - Ámbito nacional
 - Ámbito autonómico
 - Ámbito local
- El tercero es un grupo en el que se clasifican los términos por una “Tipología” requerida por la empresa
- Nombres
 - Verbos
 - Entidad
- El cuarto grupo es una etiqueta denominada “Periodo de tiempo” bajo la que se clasifican los periodos de tiempo incluidos en el tesoro con el símbolo de la almohadilla para que posteriormente el sistema automático los pueda controlar

Para ilustrar esta clasificación, en la figura 2, el término de *Aguirre, Esperanza*, presenta tres de estos cuatro grupos genéricos, como entidad, política nacional y ámbito nacional. También está la cuarta clasificación, pero en ésta, el término no está directamente en ella, sino que bajo este término genérico se clasifican los periodos de tiempo en que Esperanza Aguirre ha sido Ministra de Educación, Cultura y Deporte de España, Presidenta de la Comunidad de Madrid y Presidenta del Senado de España.

<p>Aguirre, Esperanza</p> <p>TR: Ministra de Educación, Cultura y Deporte de España</p> <p>TR: Ministra de Educación, Cultura y Deporte de España #1996-1999</p> <p>TR: Presidenta de Comunidad Autónoma</p> <p>TR: Presidenta de la Comunidad de Madrid</p> <p>TR: Presidenta de la Comunidad de Madrid #2003-2012</p> <p>TR: Presidenta del Senado de España</p> <p>TR: Presidenta del Senado de España #1999-2002</p> <p>TG: Entidad</p> <p>TG: Política nacional</p> <p>TG: Ámbito nacional</p> <p>UP: Aguirre</p> <p>UP: Aguirre Gil de Biedma</p> <p>UP: Esperanza Aguirre</p> <p>UP: Esperanza Aguirre y Gil de Biedma</p>

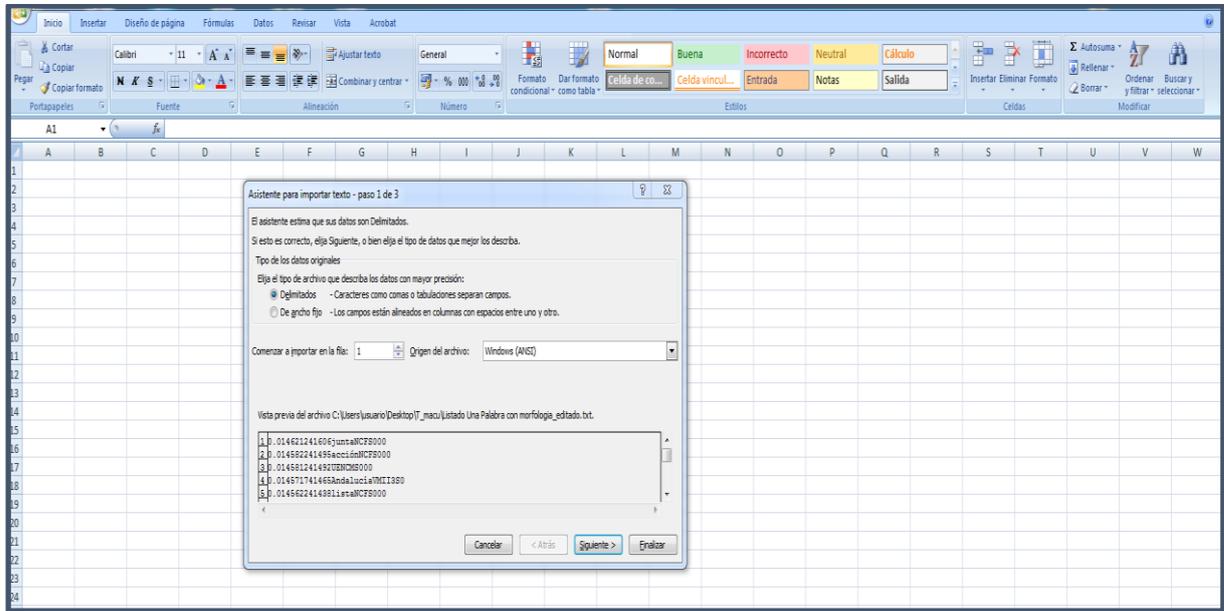
Figura 2. Presentación del término Aguirre, Esperanza

2.4 Creación

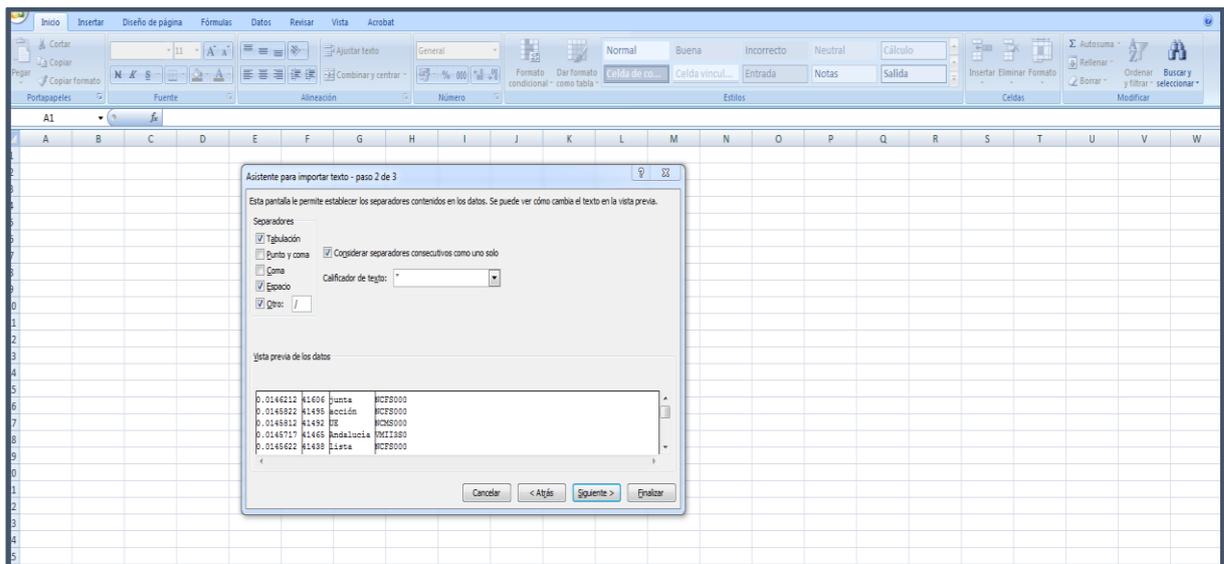
En este punto ya se sabe lo que quiere la empresa, la herramienta con la que realizar el trabajo y se cuenta también, facilitadas por la empresa, con cuatro listas de palabras reunidas tras la captura de los informativos nacionales durante un año y el contenido de las noticias de 100.000 diarios.

Se trata de cuatro listas, con expresiones de uno, dos, tres y cuatro términos, a los que se le han asignados pesos específicos y el número de veces que se han utilizado. El tamaño de las listas es considerable, todas tienen más de un millón de elementos, pero tras su revisión se opta por trabajar prioritariamente con la lista de dos términos porque es la que más información con sentido ofrece, aunque se introducen términos en el tesoro de una sola palabra y de más de dos.

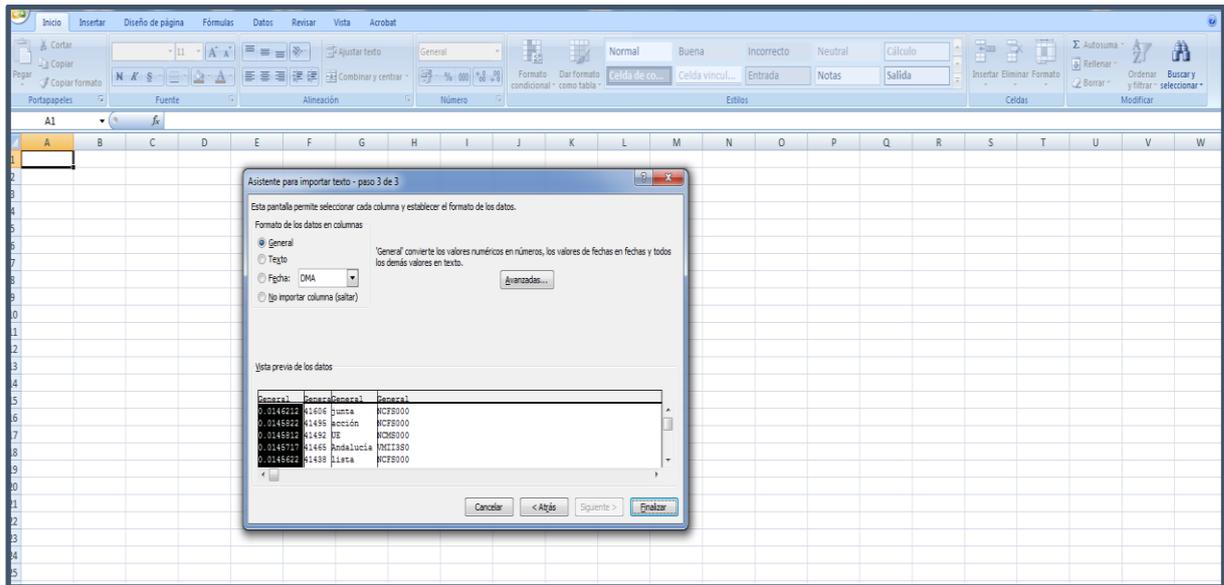
El documento que contiene la lista de palabras se debe abrir con el software Excel y utilizar el asistente para la importación de texto (captura 1). Se trata de texto delimitado con tabulaciones, por espacio y el símbolo de barra (captura 2), con formato en Excel de carácter general (captura 3).



Captura 1. Asistente para importar texto en Excel (paso 1)



Captura 2. Asistente para importar texto en Excel (paso 2)



Captura 3. Asistente para importar texto en Excel (paso 3)

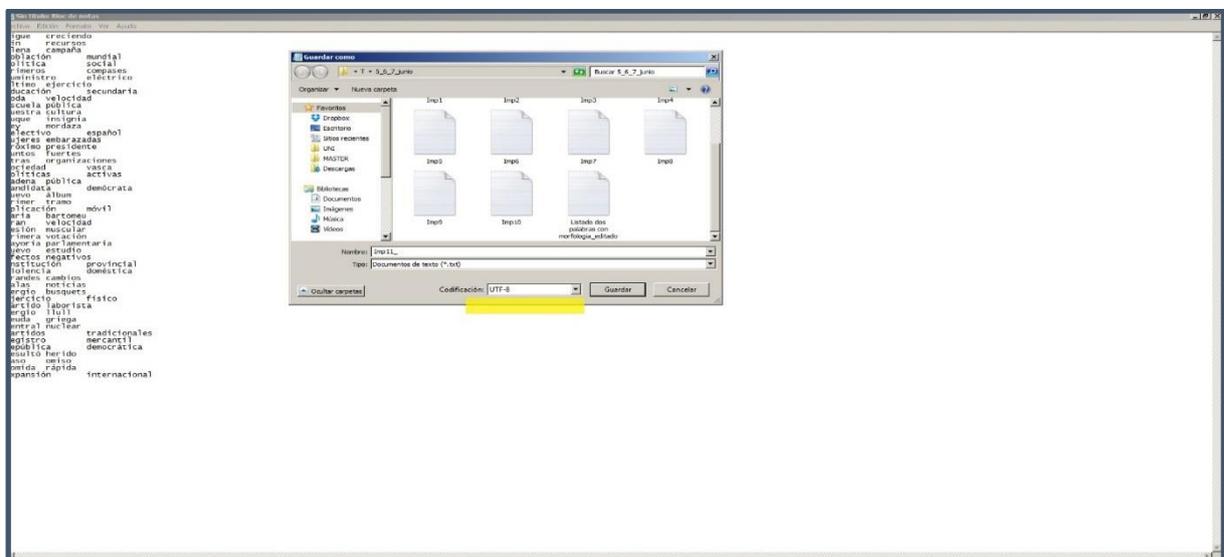
El documento una vez abierto muestra cuatro columnas, en la lista de una palabra, seis en la lista de dos (captura 4 y 5). En la captura 4 la columna A, muestra el peso del término, la columna B, el número de veces que esa palabra ha sido utilizada en el material analizado, por ejemplo, la palabra *junta* se ha utilizado 41.606 veces. En la columna D se describe el término, la palabra *junta* es un nombre (N), común (C), femenino (F) y singular (S).

	A	B	C	D	E	F	G	H	I	J
1	0.0146212	41606 junta		NCF5000						
2	0.0145822	41495 acción		NCF5000						
3	0.0145812	41492 UE		NCMS000						
4	0.0145717	41465 Andalucía		VMII3S0						
5	0.0145622	41438 lista		NCF5000						
6	0.0145615	41436 declaración:		NCFP000						
7	0.0145559	41420 China		AQQFS0						
8	0.0145105	41291 hijo		NCMS000						
9	0.0145	41261 negocio		NCMS000						
10	0.014436	41079 edición		NCF5000						
11	0.0143541	40846 Pablo		NCMS000						
12	0.014339	40803 norte		NCMS000						
13	0.0142912	40667 Francia		NCF5000						
14	0.0142821	40641 gol		NCMS000						
15	0.014229	40490 población		NCF5000						
16	0.0142171	40456 libertad		NCF5000						
17	0.0141812	40354 superior		AQQCS0						
18	0.0141595	40292 ingresos		NCMP000						
19	0.0141559	40282 fuerzas		NCFP000						

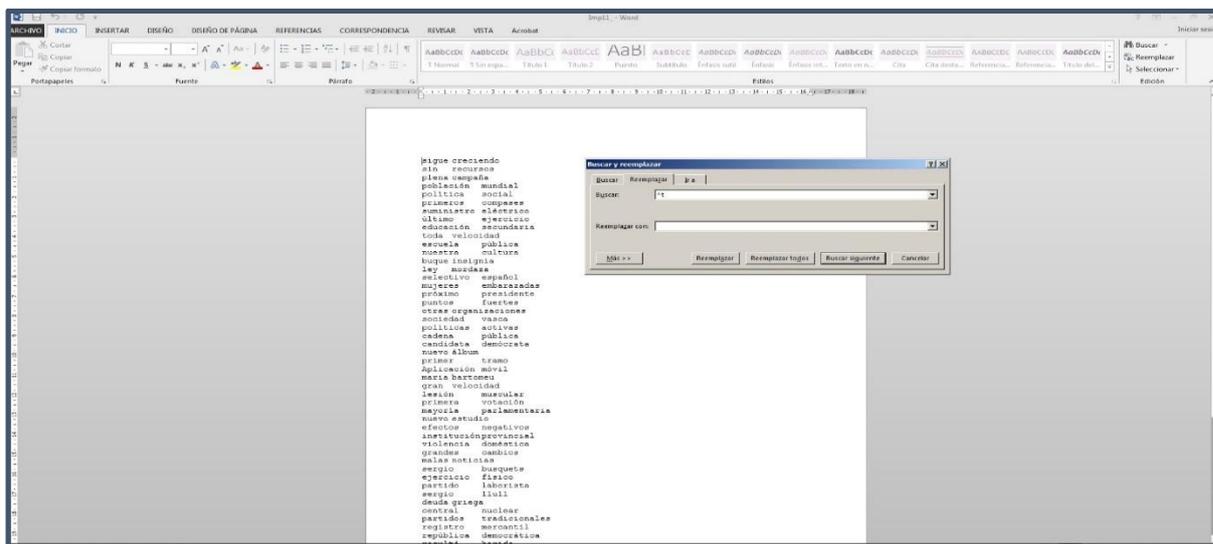
Captura 4. Lista de una palabra abierta en Excel

Captura 5. Lista de dos palabras abierta en Excel

Para editar el texto se procede a copiar a un documento de texto sin formato, únicamente las palabras (captura 6), y guardarlas con codificación UTF-8 (captura 6), necesaria para la correcta importación desde TemaTres. Aunque en el caso de la lista de dos palabras todavía es necesario sustituir la tabulación existente en el documento de origen por un espacio, ya que si no al importar desde TemaTres el archivo con los términos lo interpreta como si fuera una sola palabra (captura 7).

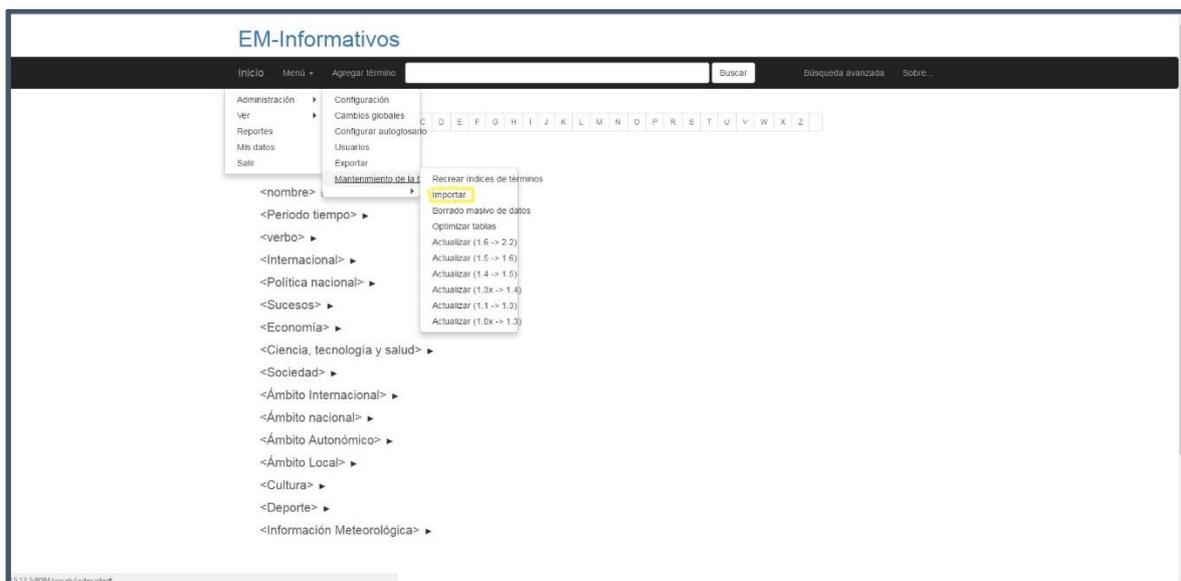


Captura 6. Edición de las palabras y codificación en UTF-8

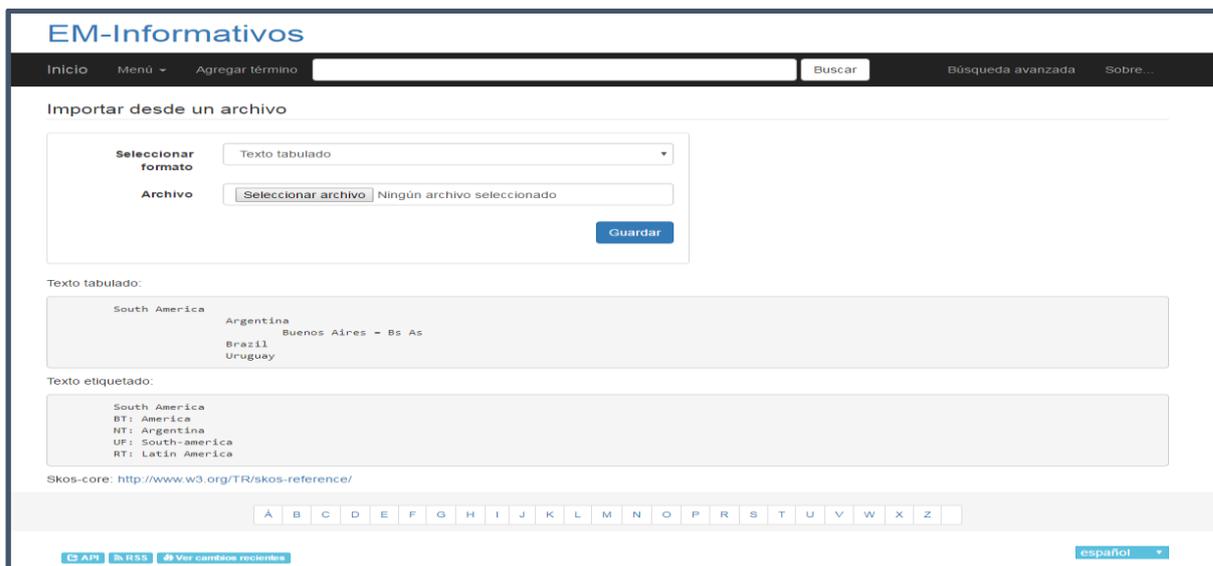


Captura 7. Sustitución de la tabulación por un espacio en la edición de la lista de dos palabras

En el proceso de importación de los términos ya editados se decide que el número de términos que se importen cada vez, sea mínimo, por la laboriosidad que implica la edición en TemaTres, por lo que se opta por no sobrepasar los cincuenta elementos en cada documento de texto e importar los términos al tesoro paso a paso. Desde la aplicación TemaTres se importa el documento de texto guardado en el paso anterior con la codificación adecuada (captura 8 y 9).

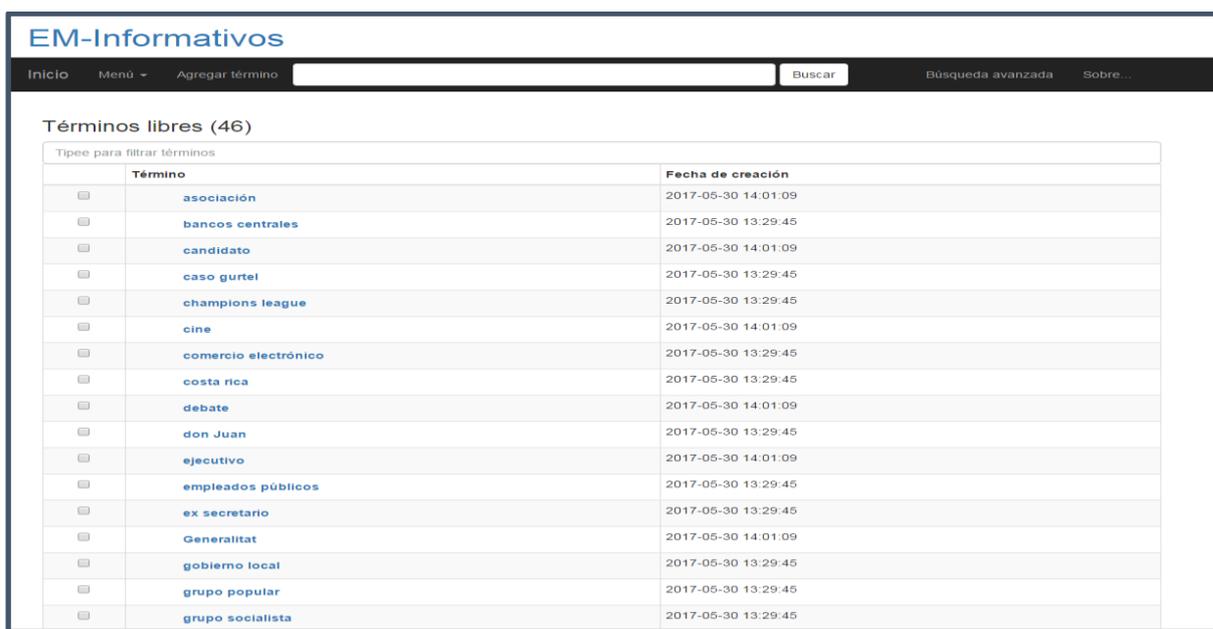


Captura 8. Importación de documentos desde TemaTres



Captura 9. Selección del archivo a importar en TemaTres

El contenido del documento de texto importado, aparece en la aplicación como términos libres (captura 10), éstos deben editarse y relacionarse con otros términos, para definitivamente formar parte del tesouro.



Captura 10. Términos libres tras la importación del documento de texto

A la entrega del archivo final a la empresa, una vez acabado el periodo de prácticas, el tesouro constaba de 3.308 términos (captura 11). Esta herramienta es, con este

número de elementos, operativa para ser utilizada con la aplicación y realizar las pruebas oportunas para llevar a término esta memoria.

The screenshot shows the EM-Informativos / ETIQMEDIA website. The header includes navigation links like 'Inicio', 'Menú', 'Agregar término', and a search bar. The main content area displays metadata for the 'ETIQMEDIA' vocabulary, including URI, language, creation/modification dates, and a list of terms categorized by depth level. A table titled 'Auditoria de términos' provides a monthly breakdown of term counts for the years 2017.

Términos según nivel de profundidad		
Nivel	# de términos	
Nivel 1	20	
Nivel 2	2499	

Auditoria de términos		
Año	Mes	Términos
2017	Jun	1947
	May	1361
Total de términos:		3308

Captura 11. Datos del tesoro a la entrega del documento final

Las normas para la edición del tesoro se establecen de acuerdo a las necesidades requeridas por la empresa, pero manteniendo la formalidad establecida para el control del vocabulario.

Las entradas de los nombres de las entidades se realizan, de forma inversa en el caso de las personas con el apellido en primer lugar (figura 3) y con nombres oficiales en el caso de las empresas e instituciones (figura 4), los lugares (figura 5) y los cargos (figura 6).

Alaya Rodríguez, Mercedes Carmen
TG: Entidad
TG: Política nacional
TG: Ámbito nacional
UP: Jueza Alaya
UP: Mercedes Alaya

Figura 3. Descriptor editado de Mercedes Alaya

Corporación de Radio y Televisión Española
TG: Entidad
TG: Política nacional
TG: Sociedad
TG: Ámbito Internacional
UP: RTVE
UP: Televisión Española
UP: TVE

Figura 4. Descriptor editado de la Corporación de Radio y Televisión Española

A Coruña
TG: Entidad
TG: Ámbito Local
UP: La Coruña

Figura 5. Descriptor editado de A Coruña

Presidente de Gobierno
TR: Aznar López, José María
TR: Calvo-Sotelo, Leopoldo
TR: González Márquez, Felipe
TR: Rodríguez Zapatero, José Luis
TR: Suárez González, Adolfo
TR: Rajoy Brey, Mariano
TG: Entidad
TG: Política nacional
UP: Jefe del Ejecutivo
UP: Presidente del Gobierno

Figura 6. Descriptor del cargo de Presidente de Gobierno

En las entradas de los cargos es necesario incorporar conceptos distintos para distinguir entre géneros. Ante la búsqueda de Presidente de Andalucía, se obviaban las noticias en las que la protagonista era la Presidenta andaluza Susana Díaz, por lo tanto se opta por crear los cargos con distinción de género en todos los cargos incorporados al tesoro (figura 7).

<p>Presidenta de la Junta de Andalucía</p> <p>TR: Díaz, Susana</p> <p>TG: Política nacional</p> <p>TG: Ámbito Autonómico</p> <p>UP: Presidenta andaluza</p> <p>UP: Presidenta de Andalucía</p> <p>Presidente de la Junta de Andalucía</p> <p>TR: Chaves González, Manuel María</p> <p>TG: Política nacional</p> <p>TG: Ámbito Autonómico</p> <p>UP: Presidente andaluz</p> <p>UP: Presidente de Andalucía</p>

Figura 7. Descriptores editados de los cargos de la Presidencia de la Junta de Andalucía

Las entradas de los nombres comunes se realizan manteniendo la ortografía de la lengua española (figura 8), en minúsculas y utilizando como descriptor su forma en singular, supeditando los plurales a términos equivalentes (estos son a los que se le asigna la etiqueta UP). Los verbos se presentan en el tesoro en su forma en infinitivo (figura 9).

<p>autoridad</p> <p>TG: Nombre</p> <p>TG: Política nacional</p> <p>TG: Ámbito Internacional</p> <p>UP: autoridades</p>

Figura 8. Descriptor editado de autoridad

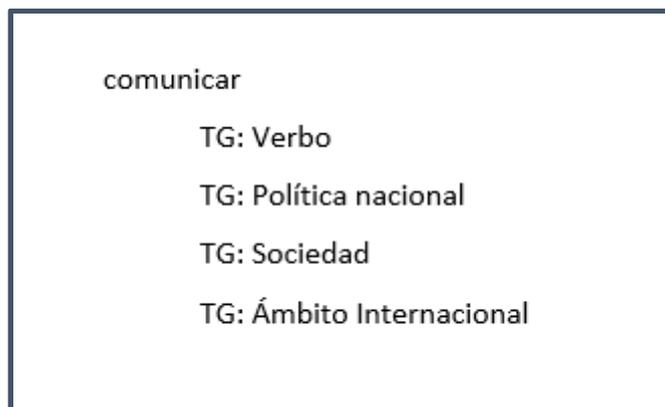


Figura 9. Descriptor editado de comunicar

Todos los términos deben estar en una clasificación como mínimo, aunque una parte importante de los elementos que forman el tesoro lo están en dos de las clasificaciones genéricas, en el gráfico 1 se presenta el número de términos por grupo clasificadorio.

En el grupo “Temática de informativos”, el número de elementos es 2.591, es decir, un 78% de los términos está indizado en esta categoría. En relación con el “Ámbito” el total de conceptos es 2.221, lo que supone un 67% del total. Son las dos categorías más numerosas, ya que son, las dos en las que se incluyen la mayoría de los descriptores. En el grupo “Tipología” hay indizados el 24% y en “Periodo de tiempo” un 3%.

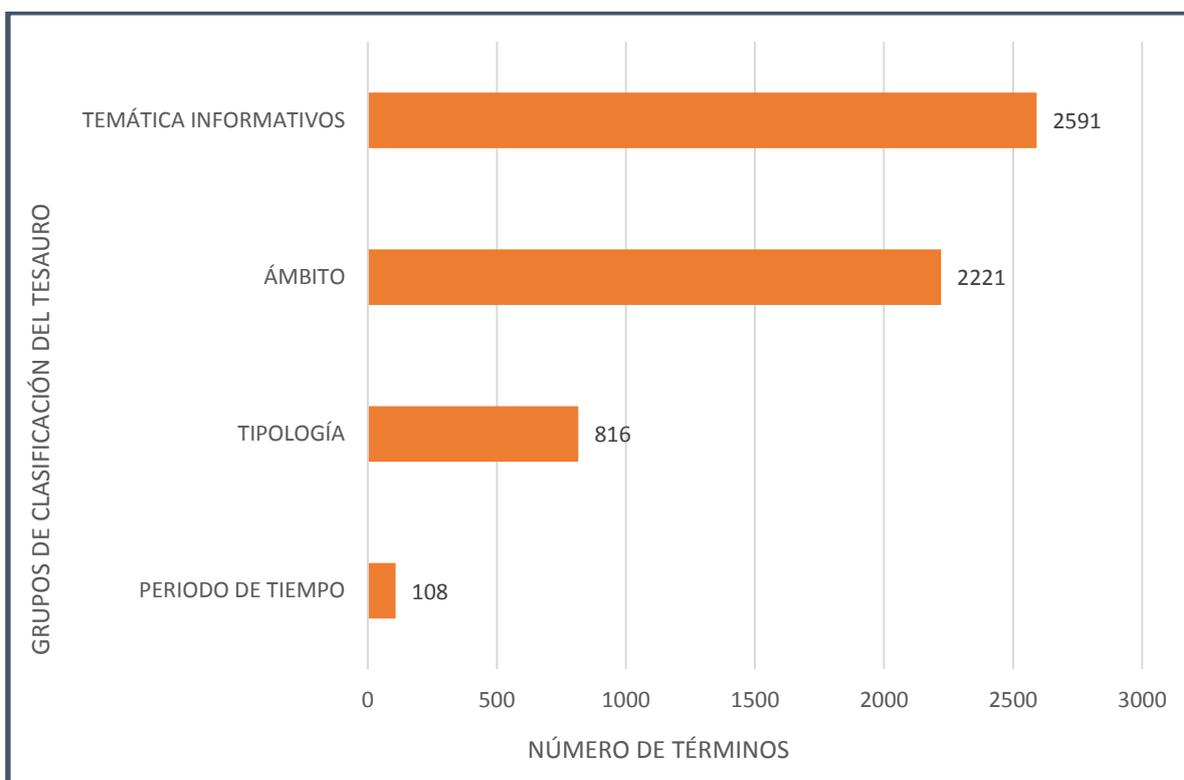


Gráfico 1. Número de términos por grupo clasificatorio

El grupo que clasifica los términos en la “Temática de informativos” se divide como se ha visto en el punto de organización del tesauro en secciones, concretamente en nueve, que se presentan detalladas cuantitativamente en el gráfico 2. La sección más numerosa en cuanto términos clasificados en ella es *Política nacional*, hay clasificados 1.133 descriptores. *Sociedad* y *Economía* tienen valores similares, con 362 y 348 elementos respectivamente. Los términos sobre *Deporte* son 253, 157 en la sección *Internacional*, 122 sobre *Ciencia, tecnología y salud* y 113 sobre *Cultura*, por debajo de 100 están los *Sucesos* con 76 descriptores y la *Información meteorológica* con 27.

En el gráfico 3 se presenta la información sobre el número de términos clasificados en el grupo “Ámbito”, que se asigna por la dimensión geográfica que puede adquirir el descriptor en una noticia. El más numeroso es el *Ámbito Internacional* con 1.432, seguido del *Ámbito Nacional* con 505, menos cuantiosos son el *Ámbito Autonómico* y el *Ámbito Local*.

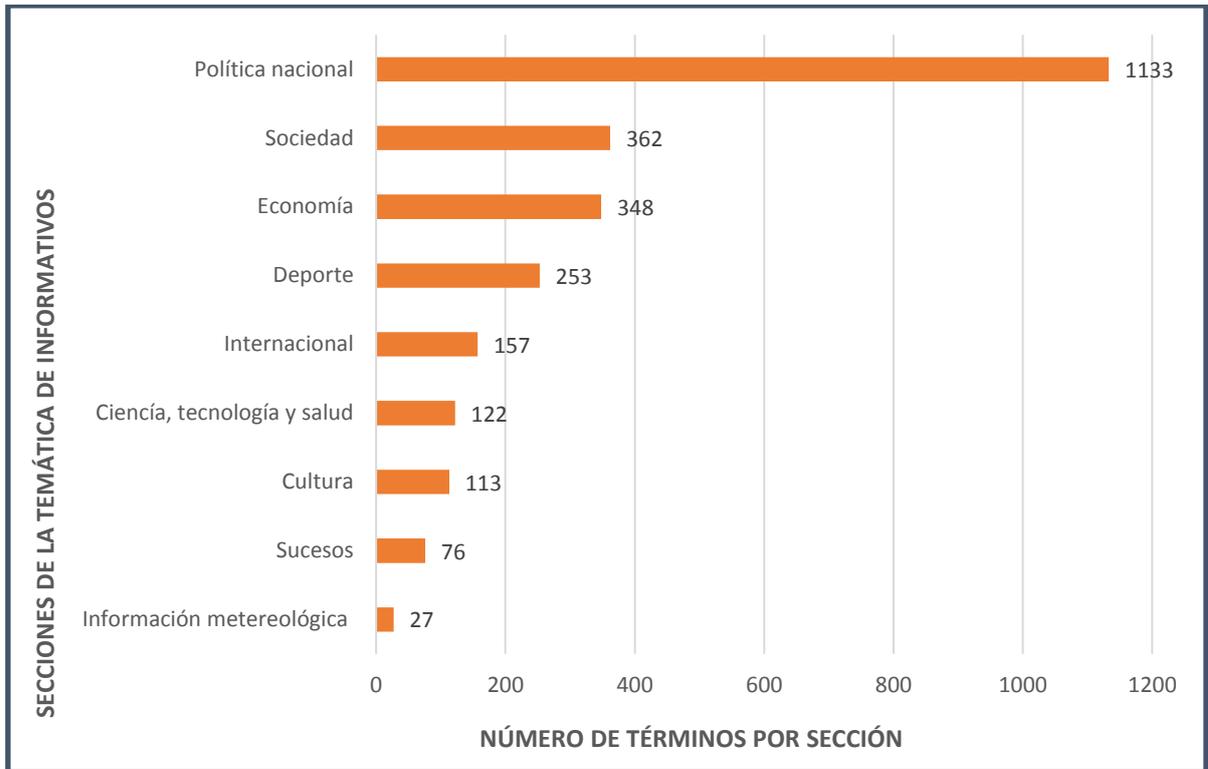


Gráfico 2. Número de términos por sección de "Temática de informativos"

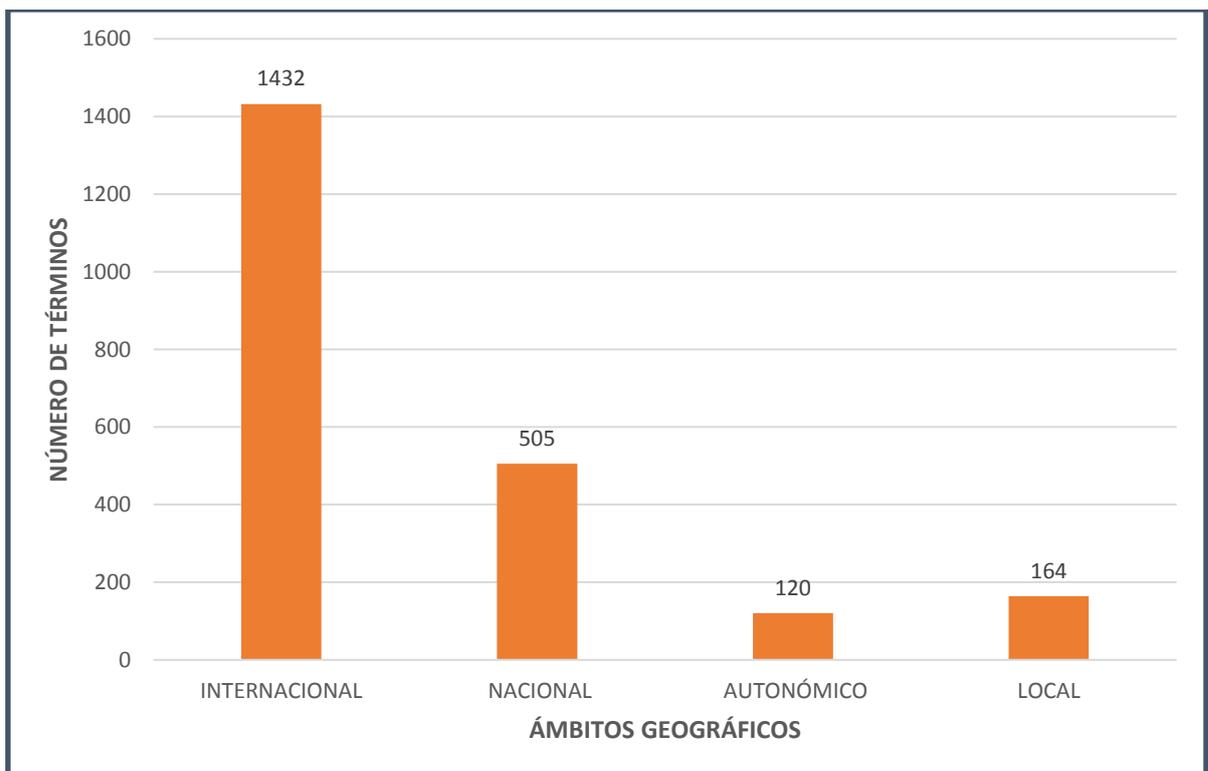


Gráfico 3. Número de términos por "Ámbito" geográfico

Queda describir cuantitativamente la clasificación de “Tipología”, representada en el gráfico 4. Lo más significativo en este grupo de clasificación son los descriptores de *Entidad*, bajo este epígrafe se clasifican las personas, los lugares, además de las empresas e instituciones.

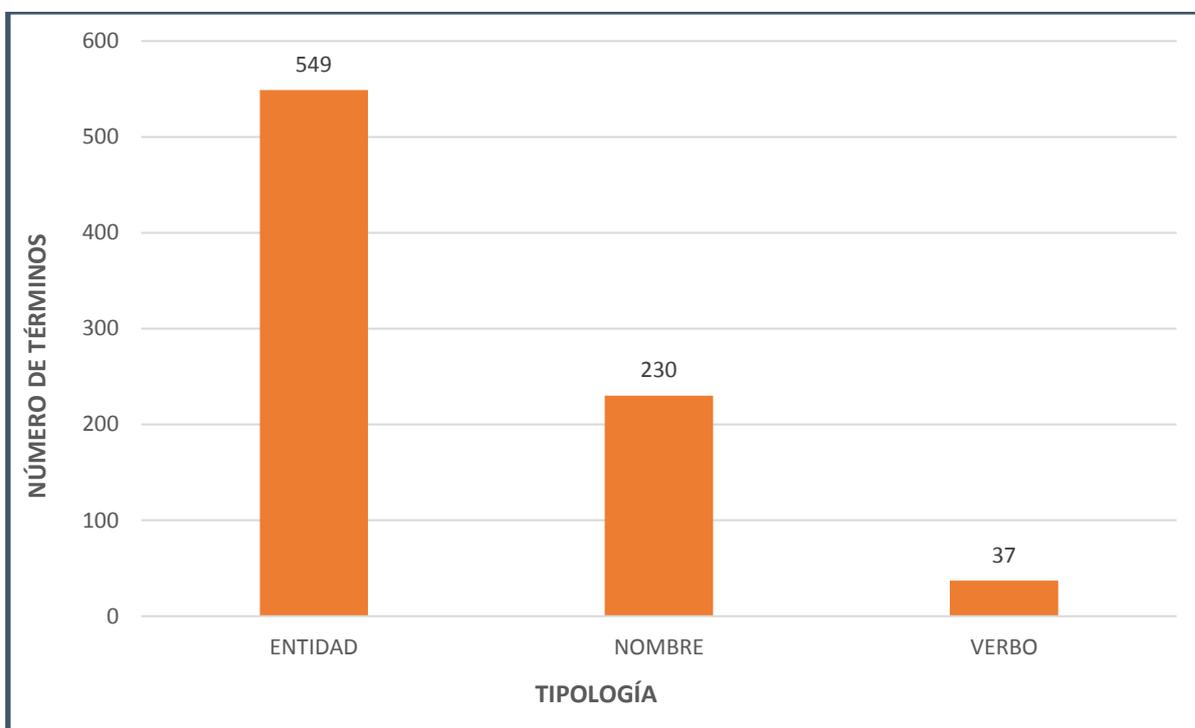


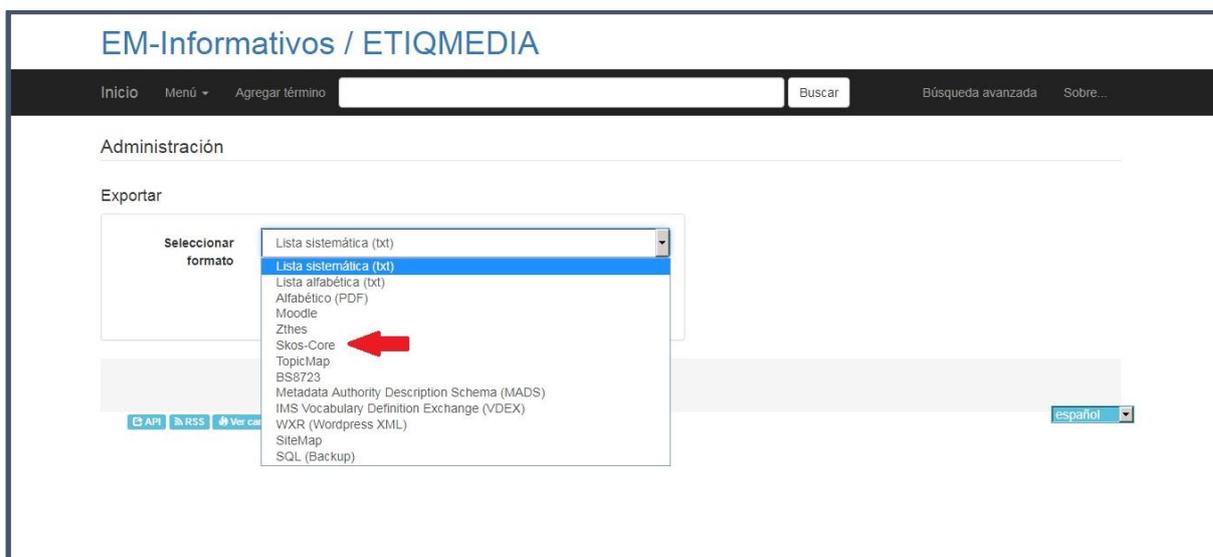
Gráfico 4. Número de términos por “Tipología”

2.5 Aplicación

De las opciones de exportación que ofrece TemaTres, se opta por Skos-core (captura 12), por lo que la aplicación genera un documento RDF¹⁸, que consiste en una familia de especificaciones del W3C, originalmente diseñado como un modelo de datos para metadatos. Es un método general para la descripción conceptual o modelado de la información orientado a recursos web, utiliza una gran variedad de notaciones de sintaxis y formatos de serialización de datos y por ello ha sido el elegido por la

¹⁸ Del inglés Resource Description Framework. Fuente *Wikipedia*. Disponible en: <http://tinyurl.com/ycdf6bg6>

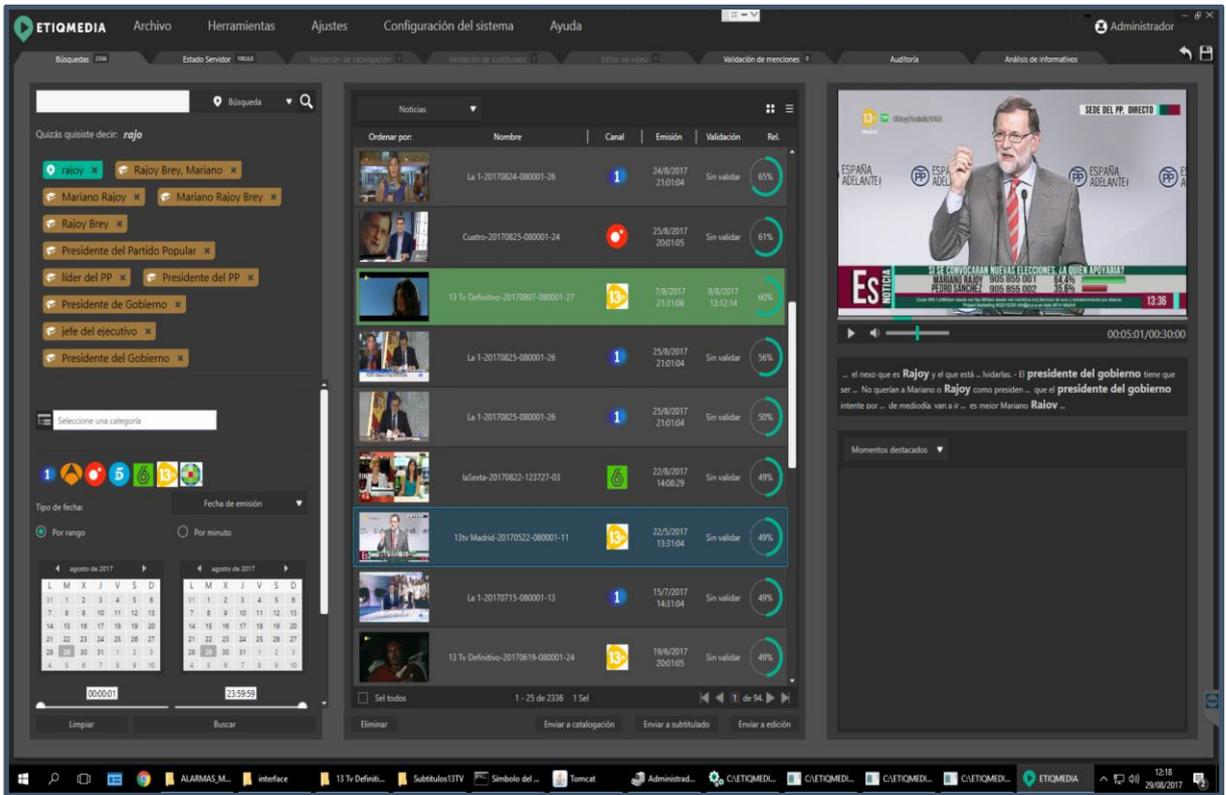
empresa, gracias a su interoperabilidad con la aplicación implantada para la clasificación de los contenidos audiovisuales.



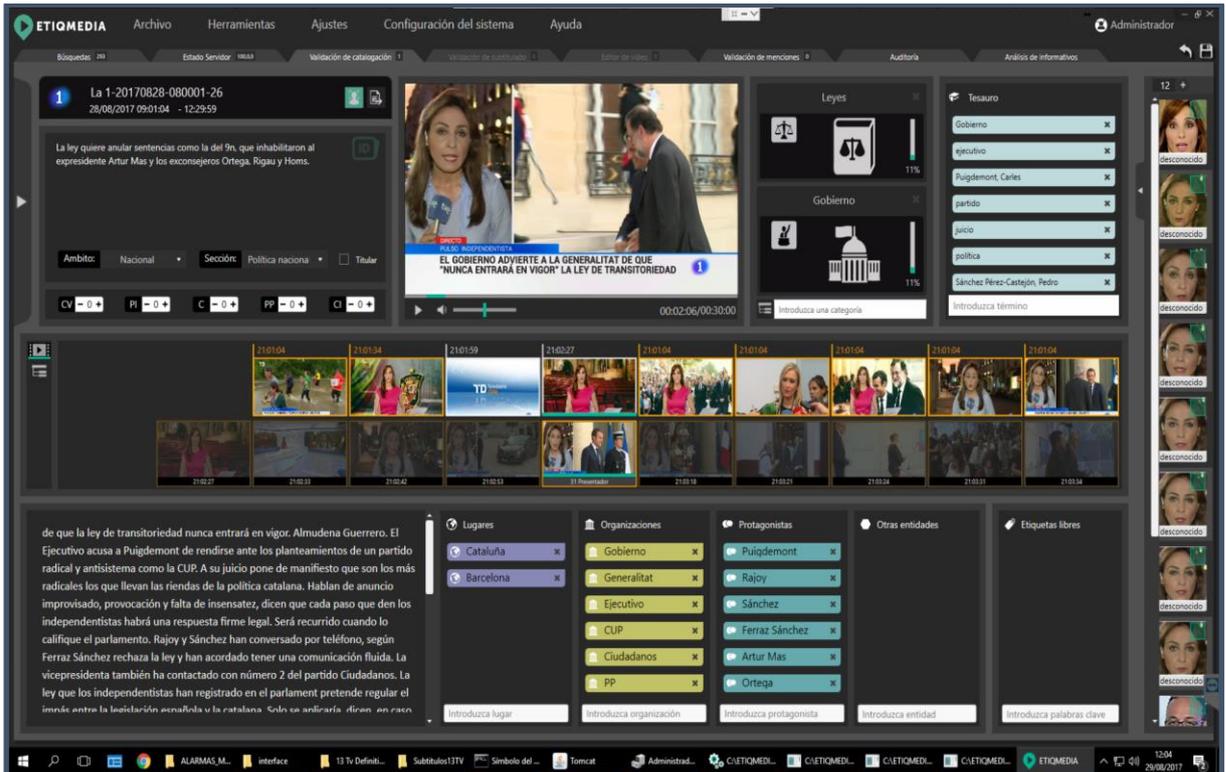
Captura 12. Opciones de exportación del tesoro

Cuando se utiliza el tesoro con la aplicación de ETIQMEDIA se realizan las funciones listadas anteriormente:

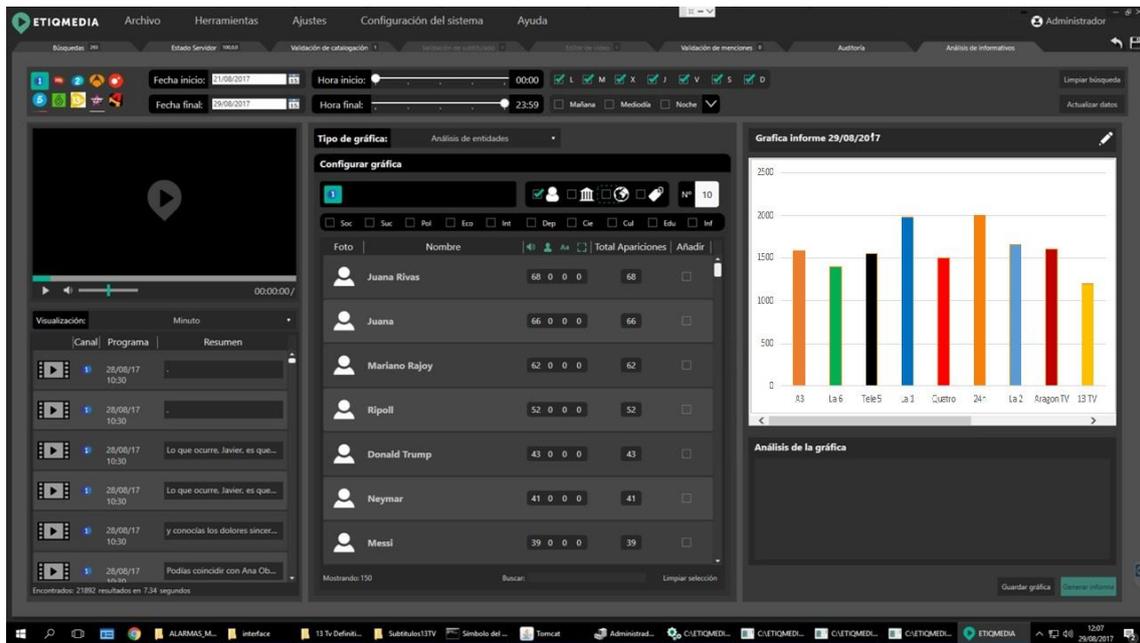
- Reunir bajo un solo descriptor todas las posibles formas de nombrar a una misma entidad y facilitar la búsqueda en los contenidos clasificados por medio de los descriptores y los términos asociados a ellos (captura 13 y 14)
- Servir como herramienta para analizar los contenidos de las noticias y realizar análisis y estadísticas (captura 15)



Captura 13. Búsqueda del término “rajoy”



Captura 14. Términos del tesoro incluidos en la noticia seleccionada



Captura 15: Análisis y estadísticas

3 Conclusiones

Desde el inicio la Documentación encuentra sus bases científicas en la Teoría General de la Información y Comunicación. Y teóricos como Otlet en 1934 señalaron a comienzos del siglo XX que la Documentación era una rama de una clasificación más general: los medios de información y comunicación.

La irrupción de las nuevas Tecnologías de la Información y Comunicación (en adelante TIC) en la sociedad en general, y en la sociedad del conocimiento en particular, ha influido y modificado, entre otras cosas, el proceso de RI con todo lo que conlleva. Al tradicional binomio poder-información se le ha sumado un tercer elemento, la documentación, que viene a reforzarlo y consolidarlo, aportando más información sobre los datos.

Sin embargo, y a pesar de las ventajas que ofrece la WWW, en materia de RI a través de los tesauros, todavía no se está aprovechando toda su capacidad, bien por falta de interés de las empresas de comunicación y/o bien por falta de recursos económicos.

En lo que respecta al tesouro para informativos de televisión, desarrollado en este TFM, y a la finalización de las prácticas, se ha puesto a disposición de la empresa una herramienta con un total de 3.308 términos, cifra que cubre hasta el momento todas las necesidades informativas de ETIQMEDIA y que está en consonancia con las de tesauros similares consultados.

Por un lado, entre las principales conclusiones destacar que:

- Una vez analizadas otras alternativas existentes en el mercado como pueden ser las listas de términos, las taxonomías y las ontologías, se dedujo que lo que necesitaba la empresa era un tesouro propio, ya que es la solución que proporciona más posibilidades de cara a la indización, navegación y futura RI
- La disponibilidad de un tesouro propio de informativos es una herramienta de gran valor documental y de comunicación para la Institución o empresa que lo posea

- El tener un tesoro propio facilita el trabajo de la importante tarea del establecimiento de relaciones de sinonimia así como la clasificación de los términos por secciones, que permitan clasificar cada minuto de un informativo, proporcionando así, un valor añadido a la información del que hasta ahora se carecía.
- Las nuevas TIC están facilitando y haciendo posible extraer mucha más información de la que tenemos ahora a nuestro alcance, reforzando así el histórico tándem poder-información gracias a disponer de una documentación más organizada y accesible
- Los medios de comunicación deben entender que la creación de un tesoro no es un fin en sí mismo sino un medio para acceder con rapidez y eficacia a las fuentes de información analizadas que necesitan una constante revisión y actualización

Y por otro lado, nos gustaría comentar algunas líneas de trabajo futuro que hacen factible la continuidad de un trabajo documental tan necesario e interesante como éste:

- (i) Sería conveniente que las empresas de comunicación tomaran conciencia de la importancia de los tesauros para ordenar y recuperar información, y dedicaran un mayor número de recursos humanos, económicos y materiales para el mantenimiento diario de esta herramienta con el objetivo de contar con información de actualidad, en tiempo real, dada la velocidad informativa a la que están sometidos los medios de comunicación,
- (ii) Asimismo, se podría favorecer la realización de una exhaustiva investigación sobre los tesauros de los medios de comunicación y su importancia, que actualice lo realizado hasta ahora. En esta misma línea, se podría intentar comprometer a las Instituciones para que se impliquen en la realización de tesauros, en los diferentes entes públicos de radio y televisión, a nivel nacional, mediante becas y convenios con las universidades

- (iii) Dado que las últimas investigaciones en materia de creación de tesauros, resaltan la importancia de contemplar las necesidades de búsqueda de los usuarios, sería conveniente profundizar en el diseño de un modelo de relaciones más apropiado, al objeto de poder resolver en menos tiempo y con mayor fiabilidad, los requerimientos informativos de éstos.

En definitiva y para finalizar con este apartado, creemos que la realización de tesauros vinculados a la actualidad informativa es un trabajo que se debería potenciar e incrementar, dada la importancia de esta herramienta en materia de RI documental.

4 Licencia documental



5 Bibliografía

- AENOR (2011). *UNE-ISO 25964 Información y documentación. Tesoros e interoperabilidad con otros vocabularios. Parte 1: Tesoros para la recuperación de la información*. Madrid: AENOR.
- AENOR (1990). *Asociación Española de Normalización y Certificación. Directrices para el establecimiento y desarrollo de tesauros monolingües: Norma UNE 50-106-90. Equivalente a ISO 2788-1986*. Madrid: AENOR.
- Agirreazaldegui Berriozabal, T. (1996). *El uso de la documentación audiovisual en los programas informativos diarios de televisión*. San Sebastián: Universidad del País Vasco, Facultad de Ciencias Sociales y de la Comunicación, Departamento de Periodismo. Disponible en: <http://tinyurl.com/ybcavc2y> [Consulta: 19/06/2017].
- Aichison, J. y S. D. Clarke (2004). El tesoro: un punto de vista histórico, con una mirada hacia el futuro. *Catalogación y clasificación trimestral*. 37 (3-4), 5-21.
- Agustín Lacruz, C. (1996). *Lingüística documental y lenguajes documentales: notas sobre un concepto, su tipología y evolución histórica*. Zaragoza: Universidad de Zaragoza.
- Arano Poggi, S. (2005). Los tesauros y las ontologías en la Biblioteconomía y la Documentación. *Anuario Hipertext.net*. 3 mayo. Disponible en: <https://tinyurl.com/y9eyvgx8> [Consulta 14/06/2017].
- Caldera Serrano, J. (2015). Externalización de la gestión documental audiovisual en las televisiones por medio de productoras audiovisuales y agencias de noticias. *Transformação*. 27(2), 145-156. Disponible en: <http://tinyurl.com/y8k6rd5w> [Consulta: 02/08/2017].
- Castells Oliván, M. (2009). *Comunicación y poder*. Madrid: Alianza Editorial.

Caro, C. y M. San Segundo (1999). Lenguajes documentales y exclusión social. En: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información*. Actas del IV Congreso ISKO-España EOCONSID'99, 22-24 abril de 1999. Granada, 101-108.

Castillo, L. y A. de la Cueva (2007). Evolución y uso de los lenguajes controlados en documentación informativa. *El profesional de la información*. Noviembre-diciembre, 16, 6, 617-626.

Chacón Gutuiérrez, I. (1994). *La fiabilidad del proceso documental del discurso periodístico, como fuente de información: análisis, método y valoración*. Madrid. Universidad Complutense de Madrid. Facultad de Ciencias de la Información. Departamento de Periodismo III.

Chaumier, J. (1986). *Análisis y Lenguajes Documentales*. El tratamiento lingüístico de la información documental. Barcelona: Mitre.

Codina, Ll. y M. E. Fuentes i Pujol (1999). Documentación periodística y bases de datos: elemento para su fundamentación como disciplina y propuesta de conjunto nuclear de diccionario de datos. *Bibliodoc: anuari de biblioteconomia, documentació i informació*. Disponible en: <http://tinyurl.com/y74733l2> [Consulta: 05/09/2017].

Currás Puente, E. (2012). *Ontologías, taxonomías y tesauros en teoría de sistemas y sistemática*. Madrid, Nostrum.

Currás Puente, E. (2005). *Ontologías, taxonomía y tesauros. Manual de construcción y uso*. Gijón, Trea.

Del Valle, F. y A. García (2003). Tesauros e Información Audiovisual. Estudio de caso. *Documentación de las Ciencias de la Información*. 26, 165-180.

Dextre, S.G. y M. Lei (2012). *From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling*. Disponible en: <http://tinyurl.com/ybtngzdy> [Consulta: 19/06/2017].

Fernández-Pampillón Cesteros, A.M. (2010). *La construcción de tesauros académicos: un modelo general y un método inductivo con aplicación al e-learning*. Madrid: Universidad Complutense de Madrid. Disponible en: <http://tinyurl.com/yaojhkwI> [Consulta 19/06/2017].

García Gutiérrez, A. (1999). *Procedimientos de análisis documental automático*. Sevilla: Junta de Andalucía, Consejería de Cultura.

García Jiménez, A. (2000). El Centro de Documentación Informativa de Telecinco. En Galdón López, G. (coord.). *Teoría y práctica de la Documentación Informativa*. Barcelona: Ariel. 223-242.

García Marco, F. J. (2014). La nueva norma de tesauros ISO UNE 25964-1. *Anuario ThinkEPI*. 8, 159-163.

García Marco, F.J. (coord.) (2007). Proyectos internacionales de reforma y ampliación de las normas sobre tesauros para su adaptación a los nuevos contextos de integración e interoperabilidad en el entorno digital. En Rodríguez Bravo B. y M.L. Alvite Díez. *La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico*. Actas del VIII Congreso ISKO-España, León, 18,19 y 20 de abril de 2007. León: Universidad de León, Secretariado de Publicaciones. 389-398. Disponible en: <http://tinyurl.com/yae5u8sg> [Consulta 19/08/2017].

García, A. y R. Lucas (1987): *Documentación automatizada en los medios informativos*. Madrid, Paraninfo.

Gil Urdiciáin, B. (2004). *Manual de lenguajes documentales*. Gijón, TREA.

Gilchrist, A.; J. Aitchison y D. Bawden (2000). *Thesaurus construction and use: a practical manual*. 4ª ed. London, Aslib IMI.

López Alonso, M.A. (1998). Evolución histórica y tendencias observables en los tesauros. *Boletín Millares Carlo*. 17, 357-370. Disponible en: <http://tinyurl.com/yavyvh2j> [Consulta 31/08/2017].

López Yepes, J. (2000). Los investigadores como creadores de lenguaje científico. Introducción al estudio terminológico de la documentación en España. En Gonzalo, C. y V. García (eds.). *Documentación, terminología y traducción*. Madrid: Síntesis.

López Yepes, J. (2007). El nuevo profesional de la información, del conocimiento y de la comunicación. El bibliotecario universitario. *Anales de Documentación*. 10, 263-279.

Llobet, M. e I. Pañella (1988). Un thesaurus aplicat a un mitjà de comunicació audiovisual. Experiència a TV3. *Item*, 2-3, 51-60.

Martín Gavilán, C. (2009). Temas de Biblioteconomía. Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia, descriptores y tesauros. Disponible en: <http://eprints.rclis.org/14817/1/lendoc.pdf> [Consulta 14/06/2017].

Martín, J. y López, J. (1995). La documentación audiovisual en RTVE. *Documentación de las ciencias de la información*. 18, 143-171.

Martínez de Sousa, J. (1993). *Diccionario de redacción y estilo*. Madrid, Pirámide.

Mediavilla, M. y R. Folla (2014). *Tesoro de Mujeres*. Madrid, Biblioteca de Mujeres. Disponible en: <https://tinyurl.com/yaoz825c> [Consulta, 14/07/2017].

Mochón, G. y A. Sorli (2010). Tesauros en acceso abierto en Internet. Un análisis cuantitativo. *Revista Española de Documentación Científica*. 33, 4, octubre-diciembre, 643-663. Disponible en: <http://tinyurl.com/yag22d9r> [Consulta 19/06/2017]

Monroy Muñoz, M.A. et al. (2012). Organización de la información de recursos digitales audiovisuales y otras entidades de información. *Cuadernos de Documentación Multimedia*. 23, 44-51. Disponible en: <http://tinyurl.com/yb48fb4q> [Consulta 01/09/2017]

- Naumis Peña, C. (2007). *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México: Alfagrama Ediciones, Universidad Nacional Autónoma de México.
- Otlet, P. (1989). *Traité de Documentation: Le livre sur le livre. Théorie et pratique*. Liège, C.L.P.C.F. (VII-Annexes) (Reimpresión)
- Otlet, P. y L. Wouters (1934). *Traité de Documentation. Le livre sur le livre. Théorie et pratique*. Bruxelles: Mundaneum.
- Peniche, S. (1992). *Vocabulario controlado en Bibliotecología, Ciencia de la Información y temas afines*. México: Universidad Nacional Autónoma de México.
- Perpinyà Morera, R. (1995). Los lenguajes documentales. En Fuentes Pujol, E. (ed.): *Manual de documentación periodística*. Madrid: Síntesis, 111-132.
- Sagredo, F. y J. M^a Izquierdo (1983). *Concepción lógico-lingüística de la documentación*. Madrid: Ibercom.
- Sánchez García, P. (2014). *Vocabularios estructurados en la web: visualización e integración con las aplicaciones*. Disponible en: <http://tinyurl.com/y8cgv59u> [Consulta 27/07/2017].
- Sánchez García, P. (2014). *Vocabularios estructurados en la web: visualización e integración con las aplicaciones*. Salamanca: Universidad de Salamanca, Facultad de Traducción y Documentación. Disponible en: <http://tinyurl.com/y8cgv59u> [Consulta 19/06/2017].
- Soler Monreal, M^a C. (2009). *Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores*. Valencia: Universidad Politécnica de Valencia, Departamento de Comunicación Audiovisual, Documentación e Historia del Arte. Disponible en: <http://tinyurl.com/ycdshujo> [Consulta 10/07/2017].
- W3C. *El Consorcio World Wide Web. Estándares*. Disponible en: <http://tinyurl.com/ycoacdmz> [Consulta 20/07/2017].

