

**Difusión estratégica de publicaciones
seriadas en una biblioteca nacional.
Aplicación de clasificaciones semánticas.**

Proyecto de fin de carrera
Licenciatura de Documentación

de

Enric Nogués Pastor

Dirección:

Fernanda Peset Mancebo

Universidad Politécnica de Valencia
Valencia, mayo 2011

Abstract

We analyzed the functional deficiencies of document retrieval systems based on the classic documentary languages in the context of serials: subject headings and UDC. We compare this model with other prioritizing communication and document type against topics by implementing new genre / form headings.

Descriptors

serials / document retrieval / classification / topic / genre / form

Resumen

Se analiza, en el contexto de las publicaciones seriadas, las deficiencias en el comportamiento de los sistemas de recuperación de documentos basados en los lenguajes documentales clásicos: encabezamientos de materia y CDU. Se compara este modelo con otro que prioriza la función comunicativa y el tipo de documento a su materia mediante la implementación de los nuevos encabezamientos género/forma.

Descriptores

Publicaciones seriadas / sistemas de recuperación de documentos / clasificación / materias / género / forma

Referencia de citación recomendada

NOGUÉS PASTOR, Enric. *Difusión estratégica de publicaciones seriadas en una biblioteca nacional. Aplicación de clasificaciones semánticas*. Directora: Fernanda Peset Mancebo. Proyecto fin de carrera Licenciatura de Documentación. Universidad Politécnica de Valencia , 2011

Sumario

1. Introducción	p. 5
2. Objetivos y metodología	
1. Determinación de la hipótesis a verificar	p. 9
2. Diseño de la prueba	p. 9
3. Estado de la cuestión	
1. "Infoxicación" y sobrecarga informativa	p. 13
2. La ambigüedad del lenguaje	p. 16
3. Avances recientes en recuperación de información	p. 18
4. El vuelco lingüístico de las clasificaciones bibliográficas	p. 20
1. Comunicación versus contenido	p. 23
2. Descriptores tipológicos y descriptores de materia	p. 25
5. La recuperación de las publicaciones seriadas	p. 29
4. Realización de la prueba	p. 35
1. Análisis y descripción de modelos	p. 35
2. Análisis de descriptores según su ambigüedad	p. 40
3. Determinación de las variables de comparación de la prueba	p. 46
4. Resultados en el catálogo de la BNE	p. 48
5. Resultados en el catálogo de la BV	p. 51
6. Resultados de la prueba	p. 53
5. Conclusiones	p. 59
6. Referencias	p. 63
7. Índices de figuras y tablas	p. 67

1. Introducción

La comunicación eficaz de la información es un elemento de fundamental importancia para el bienestar y el desarrollo en cualquier sociedad (Guinchat, 1992). Por ello los gobiernos y algunas organizaciones internacionales¹ se ocupan de la misma permanentemente. Además, desde un prisma democrático, están asociados a la misma aspectos políticos tan sensibles como la promoción de igualdad de oportunidades mediante el acceso a la información sin discriminación (España, Constitución 1978).

En el contexto actual uno de los obstáculos que dificultan esta difusión es la creciente superabundancia informativa. A este fenómeno se le ha denominado "infoxicación" (Cornella, 2000) y también desinformación por exceso de información². La "infoxicación" o sobrecarga informativa dificulta la identificación de la información relevante y con ello la gestión eficiente de toda actividad, sea personal, profesional o social.

Resolver los problemas relacionados con la gestión de la información contenida en los documentos ha sido una constante desde la antigüedad. A ello se han dedicado junto con bibliotecarios y archiveros, grandes filósofos y científicos. Si bien, en cada época histórica los retos han sido distintos, su esencia, el establecimiento de caminos entre el saber contenido en los documentos y los seres humanos que los tenían que recorrer, ha existido siempre. Listas de autores, de asuntos, de títulos, etc. primero. Pronto se hicieron necesarias clasificaciones, para disponer los documentos en un cierto orden, y normas para organizar aquellas listas conforme a éstas. Las grandes clasificaciones enciclopédicas surgieron de los grandes

¹ Actualmente el Programa de Información para todos (IFAP) que ha sustituido al Programa de Información General PGI de la UNESCO

² Bajo este término se ha traducido y popularizado en España el término anglosajón Information overload acuñado por Alvin Toffler (1970).

sistemas filosóficos. El conocimiento humano recurría a la razón humana para organizarse, y gracias a ello, el volumen y complejidad de nuestro conocimiento no dejó de crecer. Después, la imprenta y las sociedades liberales multiplicaron la producción de documentos facilitando al mismo tiempo su acceso por millones de personas en la nueva sociedad de masas. Ejércitos de bibliotecarios y archiveros organizaron el saber contenido en los documentos de naciones y naciones. Surgieron las bibliotecas nacionales, las bibliografías nacionales, los sistemas bibliotecarios y las organizaciones bibliotecarias internacionales. Se soñó (La Fontaine, 1895) con el control bibliográfico universal y con el acceso universal al documento...

Sin embargo, mucho antes de alcanzar los objetivos soñados éstos se alejaban. Una nueva revolución sustituía la *galaxia Gutenberg* de lo impreso (McLuhan, 1998) por otro universo mil veces más vasto, rico y complejo: el virtual, con Internet y las comunicaciones inmediatas; el de la ubicuidad de la información, ahora liberada de la necesidad de un soporte físico, o al menos de uno en su forma clásica. Las grandes clasificaciones se quedaron pequeñas ante la inmensidad del reto. Dewey y Otlet, los procedimientos bibliotecarios y las bibliotecas mismas perdieron su papel en primera línea ante el nuevo reto de la organización del conocimiento. Se buscaron las soluciones en los ordenadores, a los que se les encargó comunicar el conocimiento y las necesidades humanas del mismo. Se hicieron estudios filológicos para ayudar a las computadoras a procesar el lenguaje humano de los documentos y de las consultas. Se hicieron, y se siguen haciendo investigaciones en el marco de la inteligencia artificial para dotar estas máquinas de más recursos inteligentes con los que abordar el formidable encargo. Al mismo tiempo desde el enfoque documental se estudiaron y desarrollaron con cierto éxito, sobre todo en áreas del saber concretas y delimitadas, lenguajes

documentales controlados.

Ya desde finales del siglo XX y en el siglo XXI muchas han sido las iniciativas puestas en marcha en distintos ámbitos y disciplinas como la Documentación, la Lingüística, la Informática o la Inteligencia artificial para solucionar este problema (Castillo, 2006). Al hilo de las mismas han surgido conceptos y propuestas nuevas y esperanzadores como los *topic maps*, las representaciones de dominios, las redes semánticas, los tesauros y clasificaciones, los mapas conceptuales o la *web* semántica (Lamarca, 2009).

El objetivo de este trabajo se inserta en este contexto. Nos planteamos, para un campo concreto, el de la gestión de las publicaciones seriadas en las hemerotecas, la evaluación comparada de dos modelos teóricos de recuperación documental. De una parte, el modelo A o clásico, basado en la clasificación/recuperación de los documentos por sus materias. De la otra, un modelo nuevo, o B, en el que, desde un enfoque sistémico centrado en el rol comunicativo de los documentos en tanto que flujo de información vital para el funcionamiento de la sociedad, y en línea con los últimos avances científicos y tecnológicos, ha sido enriquecido semánticamente mediante la aplicación de los nuevos encabezamientos género/forma (Library of Congress, 2011) que son su piedra angular.

2. Objetivos y metodología

2.1 Determinación de la hipótesis a verificar

Pretendemos mediante este estudio verificar o rechazar la hipótesis de obtención de menores niveles de ruido documental en las búsquedas realizadas a partir de sistemas de recuperación de documentos que incorporan descriptores con un contenido semántico más delimitado, menos ambiguo que exclusivamente con los clásicos encabezamientos de materia utilizados en las bibliotecas, los descriptores género/forma.

2.2 Diseño de la prueba

La metodología a utilizar consiste en la configuración de un diseño de prueba que permita evaluar de modo objetivo y estadísticamente cuantificable el comportamiento de unos y otros descriptores (variable independiente) en relación con la tasa de ruido (variable dependiente) obtenida en los documentos devueltos como resultado de las búsquedas según estén estos descriptores implementados o no en los respectivos sistemas de recuperación.

Para la realización de este diseño se han estudiado los OPAC de bibliotecas americanas, europeas y españolas, con el objeto de seleccionar de entre ellas dos sistemas, con y sin la implementación de estos descriptores, con los que confrontar la pertinencia de sus resultados. Desgraciadamente hemos comprobado que el nivel de implantación de esta herramienta es en general incipiente y que en ningún caso es posible en estos momentos la recuperación diferenciada desde los formularios de consulta de sus OPAC por este tipo de descriptores.

Se han estudiado por otra parte las principales listas de encabezamientos de materia con el objetivo de seleccionar de entre sus descriptores un grupo de encabezamientos de materia que por su carga semántica fueran especialmente susceptibles de generar ruido en las búsquedas.

Como conclusión del primer estudio y dado que no hemos encontrado ninguna biblioteca en el que el grado de implementación de estos descriptores permitiera a un usuario externo la recuperación diferenciada a partir de los mismos, hemos optado por seleccionar, en representación de los sistemas basados en recuperación por materias a la BNE porque aún habiendo iniciado la indización por este tipo de descriptores no permite en el momento actual otra recuperación que la clásica basada en las materias. Hemos seleccionado por otro lado y para enfrentar sus resultados con la anterior la BV, que aún teniendo un nivel de implantación similar sí permite la interrogación separada de sus registros a partir de este tipo de descriptores desde su formulario de consulta avanzada, y por ello nos permite comparar los resultados de consultas basadas o no en estos descriptores.

Como resultado del segundo estudio, el de las listas de encabezamientos de materia hemos podido agrupar descriptores de las mismas en función de su contenido semántico más o menos ambiguo en el contexto de las consultas de fondos de bibliotecas. Ésto nos ha permitido identificar grupos de descriptores y seleccionar para la prueba aquellos que en principio podríamos sospechar que generarían mayores niveles de ruido.

Los descriptores del grupo candidato se han utilizado en consultas en el catálogo de la BNE. Los resultados obtenidos han sido analizados determinando su pertinencia tanto con respecto a su materia como a

su tipo y a continuación han sido tabulados para obtener una visión estadística de su comportamiento.

Perteneciente al mismo grupo de descriptores se ha seleccionado uno, almanaques, para mediante la consulta diferenciada del mismo en el formulario de búsqueda de la BV obtener resultados susceptibles de comparación con los que para este descriptor se han obtenido en la BNE.

La comparación de unos y otros resultados nos permite verificar la hipótesis de partida, con la salvedad en todo caso de su carácter provisional debido a la naturaleza y los límites de un estudio como el presente. Posteriores tesis doctorales o estudios científicos podrán corroborar la corrección de la confirmación provisional de nuestra hipótesis.

3. Estado de la cuestión

3.1 "Infoxicación" y sobrecarga informativa

Guinchat (1992) afirma que el conocimiento tecnológico se ha multiplicado por diez cada 50 años desde hace más de 2800 años. Y, que a los libros y publicaciones seriadas editados por los circuitos comerciales tradicionales se suma una extensa gama de documentos de todo orden, muchos inéditos o de difusión restringida. Entre estos documentos figuran informes, tesis, actas de congresos, cursos, estudios, separatas, etc. siendo a menudo la fuente más actual y reciente de información en distintos campos y constituyendo la mejor herramienta de comunicación directa entre científicos.

Este fenómeno que Guinchat denomina *explosión documental* está vinculado con el de *information overload*, o el de "infoxicación". Éstos son términos que desde distintos ángulos y ya desde fechas tan remotas como los años treinta del siglo XX se acercan al problema de las dificultades que enfrentan los usuarios para obtener de manera eficaz y segura la información que necesitan. En la sesión inaugural del II Congreso de Bibliotecas y Bibliografía, Ortega (1935) decía:

Aún reduciendo sobremanera el número de temas a que cada hombre dedica su atención, la cantidad de libros que necesita ingerir es tan enorme que rebasa los límites de su tiempo y de su capacidad de asimilación...

Hay, pues, que crear una nueva técnica bibliográfica de un automatismo riguroso.

Toffler (1970) varias décadas después comenta las dificultades de las personas para comprender y tomar decisiones como consecuencia de un exceso de información.

Bawden y Robinson (2009) relativizan la novedad y la gravedad del problema argumentando por un lado que fenómenos similares, aunque a menor escala, se han dado en otros momentos históricos, como con la difusión de la imprenta; y por otro situándolo más como consecuencia natural de la desorientación producida por el desconocimiento y falta de familiaridad de los usuarios en los nuevos entornos en los que la información se comunica, sobre todo con las nuevas estructuras *web 2.0*: *wikis, blogs, rss...* así como por la falta de hábitos sociales y laborales adecuados. Consideran que las soluciones pueden venir por una parte de la implementación de adecuadas herramientas de filtrado selectivo de la información así como de una formación adecuada que considere, entre otros elementos, el estudio de la Historia de la información; y también el fomento de una nueva cultura que potencie la *satisfacción* de la necesidad informativa por encima de la acumulación irracional de información.

Himma (2007) analiza el concepto *information overload* desde una perspectiva de razonamiento filosófico para concluir que la solución a la sobrecarga informativa no puede estar en la disminución de la cantidad de información - que considera un bien en sí- sino en mejorar nuestra capacidad de proceso de información mediante soluciones organizativas y tecnológicas. Y, mediante la capacidad de determinación de qué información, merece compartir con otras de nuestras actividades -reflexión, meditación, diversión, participación social, etc. - la inversión de una parte de nuestro tiempo y dedicación.

Kanungo (2010) afirma que la información es necesaria para tomar

decisiones y reducir la incertidumbre, pero cuando la disponibilidad de información es superior a su demanda el usuario ha de utilizar más tiempo en comprenderla y las decisiones se retrasan.

Levy (2005) analiza el contexto y las preocupaciones que movieron a Vannevar Bush en 1945 a plantear el problema, y rescata la noción de crecimiento en sabiduría, fin último para el que es necesaria una gestión eficiente de la información. Gestión que se ha de realizar mediante los adelantos tecnológicos necesarios, como el Memex con el que en su día Bush revolucionó los planteamientos para la gestión de la información y que son la base del hipertexto, el *world wide web* e incluso de los ordenadores personales. En él, el concepto de crecimiento en sabiduría (*To grow in wisdom*) incluye la capacidad de disfrutar (*leisure*) de la vida y de los conocimientos adquiridos mediante la reflexión sosegada y la contemplación de sus resultados. Finaliza su análisis con la evidencia de que estos objetivos no se han conseguido a pesar del espectacular avance tecnológico experimentado sesenta años después de la invención del Memex.

Badke (2010) enfoca la cuestión desde la perspectiva de la alfabetización informacional (*information literacy*) subrayando que los objetivos de ésta y la formación en ellos -a saber: capacidad de identificar el problema, saber como adquirir la mejor información para solucionarlo, evaluar la información encontrada y usarla apropiadamente- tienen mucho que ver con reducir la supuesta sobrecarga informativa (*information overload*) o "infoxicación".

De esta breve revisión podemos extraer las siguientes ideas: la explosión documental no es un fenómeno nuevo, algunos autores lo sitúan desde la invención de la imprenta, aunque hay un consenso generalizado de que se vuelve más patente a partir del último tercio del siglo XIX, y de una manera acuciante tras la Segunda Guerra Mundial. La solución tecnológica es necesaria pero no suficiente, y por

ello debería venir acompañada de cambios organizativos, educativos e incluso filosófico-culturales.

3.2 La ambigüedad del lenguaje

En cualquier sistema documental el problema principal es implementar un lenguaje que establezca una comunicación eficiente entre las necesidades expresadas (o no) de los usuarios y los documentos e informaciones objeto del sistema documental dado (Maniez, 1987). Las dificultades son consecuencia de los elementos de subjetividad y ambigüedad del lenguaje.

En la comunicación humana estos problemas se resuelven gracias a que el contexto en el que la comunicación se produce aporta informaciones que objetivizan y desambiguan ese mensaje. Este fenómeno, unido a la creciente capacidad de los actuales sistemas informáticos, ha llevado durante algún tiempo a pensar que el texto completo de los documentos, procesado convenientemente por los ordenadores, facilitaría el contexto necesario para la comunicación documental. Hoy se considera (Castillo, 2006) sin embargo que las exigencias de proceso y los recursos inteligentes (reglas de inferencia, etc.) necesarias para obtener resultados aceptables desbordan con mucho las actuales capacidades informáticas. Muchos estudios avalan ahora la necesidad de combinar el uso del lenguaje natural (de los propios textos) con el documental (elaborado por especialistas) para obtener resultados aceptables.

Sabemos que los lenguajes documentales son buenos aliados para neutralizar la ambigüedad derivada de la polisemia y la sinonimia existente en el lenguaje. También, que un recurso utilizado para mejorar los resultados es reducir el ámbito de utilización del instrumento documental, es decir limitar el contexto de su utilización.

Se trata de excluir todos los conceptos que no están explícitamente declarados en el mismo. Así, se han desarrollado durante años multitud de vocabularios, glosarios y tesauros especializados. Sin embargo la multiplicidad de estos lenguajes especializados no resuelve adecuadamente las necesidades de la comunicación documental cuando ésta requiere de instrumentos generales como se da con la documentación periodística (Castillo, 2006) o la hemerográfica de la que la anterior podemos considerar como un subconjunto con su misma naturaleza.

Para enfrentarse a los problemas de "infoxicación" (o *information overload* o explosión documental) desde la perspectiva de la hemeroteca de una biblioteca cabecera de un sistema bibliotecario autonómico (Biblioteca Valenciana, 2010) tanto los instrumentos de catalogación bibliotecaria clásicos: encabezamientos de materias y CDU, como las aproximaciones al mismo realizadas desde las investigaciones mencionadas resultan insatisfactorios. Castillo (2006), estudiando el problema en el ámbito de la documentación periodística,³ expone la necesidad de continuar con las investigaciones y sitúa las dificultades que entraña el tratamiento documental del discurso informativo en su naturaleza enciclopédica en la que resultan de difícil aplicación las clasificaciones alfabéticas de materias utilizadas habitualmente en las bibliotecas generales.

Díaz (1986) considera estos instrumentos documentales pensados para la indización del conocimiento científico mientras que en la documentación de actualidad (periodística) la materia prima son acontecimientos o reflexiones sobre los mismos. Documentación que, aunque tratada a otro nivel, si se quiere con más profundidad, es una

³ Consideramos esta tesis de gran interés para nuestro trabajo dado el paralelismo y los puntos de contacto entre los problemas de la documentación periodística y los de la documentación hemerográfica, como se verá más adelante.

parte de la que gestionan las hemerotecas.⁴

Por otra parte las clasificaciones generales como la CDU además de haber sido duramente criticadas desde múltiples puntos de vista, tales como su enfoque subjetivo, el encorsetamiento al que someten a los nuevos conocimientos y su dificultad para adaptarse al ritmo actual de crecimiento de la información (Moreno, 2002), tienen también el inconveniente para nuestro propósito de estar pensadas al igual que los encabezamientos de materia de las bibliotecas desde la perspectiva y para representar prioritariamente el conocimiento científico (Díaz, 1986).

3.3 Avances recientes en recuperación de información

Históricamente la biblioteconomía/bibliografía ha aportado a la gestión de la información el enfoque, los conceptos y las herramientas con las que sus profesionales han organizado el patrimonio bibliográfico-documental.⁵ Consecuencia de ello, por ser además aceptado como parte integrante del proceso informativo global y a causa de su enorme prestigio, el enfoque bibliográfico-bibliotecario ha sido utilizado, aunque con poco éxito, para resolver las actuales necesidades informativas sociales.

La necesidad de un enfoque integrador de los distintos campos informativos: cultura (patrimonio), información (para la participación democrática), ciencia (información científica), economía y empresa (información económica) o sanidad (información médico-sanitaria) se hace más patente por las cada día más estrechas interrelaciones entre ellos. Al contrario, la especialización, también impulsada como

⁴ Las hemerotecas gestionan diarios, revistas y otras fuentes donde se recoge la información periodística.

⁵ El concepto mismo de patrimonio bibliográfico inicialmente limitado a la documentación impresa con fines de comunicación pública se ha visto con el tiempo progresivamente ampliado. Así, a libros, revistas y periódicos, se han ido añadiendo otros tipos documentales transmisores igualmente de información, conocimientos u otros valores artísticos y culturales. En la ley de Patrimonio Histórico Español (1986) se reconocen integrantes del mismo también las películas cinematográficas, discos, fotografías, materiales audiovisuales y otros similares, cualesquiera que sea su soporte material.

recurso contra la ambigüedad del lenguaje, lastra la implementación de un enfoque global que facilite políticas de coordinación de acciones a seguir en campos complementarios: investigación científica y desarrollo tecnológico, administración, cultura, etc.

Los estudios filológicos, la lógica y la computación también han contribuido, por otra parte, mediante la realización de modelos *inteligentes* para la vinculación automática de los documentos con las necesidades de los usuarios. Quizá los motores de búsqueda, con el fenómeno Google, sean su principal exponente. Tanto el anterior como Yahoo combinan en mayor o menor medida clasificaciones *bibliográficas* con motores de búsqueda basados en la creación y gestión de listas de términos (Yi, 2006).

Junto a las aproximaciones al problema descritas: insuficientes e insatisfactorias por su enfoque parcial, desmembrador del proceso informativo, y en consecuencia impotentes ante el crecimiento exponencial de la información en multiplicidad creciente de formas y soportes... también se han desarrollado en los últimos años elementos que pueden contribuir a encontrar soluciones al mismo: una visión del conjunto, desde la teoría de sistemas, y nuevos conceptos, más precisos, para su estudio basados en el análisis de los componentes del proceso informativo, su naturaleza, su objeto, su comportamiento, sus tipologías y sus formatos.

La teoría de sistemas (Moscoso, 2002) nos ha aportado la concepción del flujo informativo como una metáfora en la que la información recorre, como la sangre en un ser vivo, todo el cuerpo social aportando a cada célula los elementos vitales para su desarrollo. Una visión, en suma global, dinámica, integrada, y con leyes propias. Un sistema que permite definir la especificidad y al mismo tiempo la complementariedad en el conjunto de todas sus partes: información

no formalizada, conocimiento experiencial práctico, información débilmente formalizada, comunicaciones previas informales y formales, comunicación de información consolidada por medios con distinto grado de formalidad... vistas todas ellas como distintas fases o estadios en el proceso comunicativo. La multiplicidad de los medios existentes: escritura, sonido, audiovisual, multimedia... y de sus soportes: físicos o virtuales, analógicos o digitales. Un conjunto en suma con infinita y creciente variedad de sus partes en función de la complejidad y ambición de progreso de la sociedad que la genera.

Del lado computacional por otra parte los estudios para la clasificación automática de géneros *web* (Montesi, 2008) son también una realidad que aunque modesta, representa un ejemplo y una esperanza de una futura integración de las corrientes computacional empírico/mecánica de los motores de búsqueda con la filosófico/teórica de las clasificaciones bibliográficas.

Y finalmente, convergiendo en este enfoque, encontramos los estudios y propuestas que desde el mundo bibliotecario impulsan en la actualidad la implementación del nuevo enfoque clasificatorio género/forma en el que para los materiales especiales ya se trabaja en las principales bibliotecas (BNE, 2010).

3.4 El “vuelco lingüístico” de las clasificaciones bibliográficas

Los padres fundadores de la documentación Otlet y Lafontaine constituyeron las bases de la misma en íntima relación con el positivismo lógico de su época (Romanos, 2008). Desde esta corriente de pensamiento, el papel de las ciencias y el de la clasificación de éstas influyó poderosamente en las clasificaciones documentales. La documentación se pensó más como documentación de las ciencias que como documentación científica de las actividades

humanas, de las que la científica con ser muy importante no era la única.

Las carencias y limitaciones de este enfoque pronto relegaron a la documentación a un segundo plano en las nuevas tareas de gestión de la información (López, 2002). Sus herramientas clasificatorias clásicas (CDU, DDC, etc.) fueron sustituidas por clasificaciones u otros lenguajes documentales como los tesauros, casi siempre surgidos desde y para resolver problemas de gestión de información en sectores específicos como la informática, la medicina, la agricultura... que desarrollaron sus propios lenguajes clasificatorios al margen y desconociendo en muchos casos la teoría documental.

En el terreno filosófico el positivismo lógico y con él la fe en que la ciencia salvaría a la humanidad de la ignorancia y la superstición uniendo a la humanidad en el progreso, quedaron desautorizados por dos guerras mundiales y por la creciente brecha en el desarrollo entre los países que las sucedió. Como no podía ser de otro modo con el positivismo lógico entraron en crisis sus planteamientos y con ellos sus clasificaciones documentales ahora relegadas al mundo bibliotecario e incluso en retroceso en él.

No vamos a entrar en mayor profundidad en el devenir filosófico de las corrientes que se enfrentaron y sustituyeron al positivismo lógico más allá de resaltar una de ellas por, a nuestro parecer, el enorme interés de su punto de vista y conceptos, en relación con los grandes problemas actuales de la documentación, y sobre todo por su relación con los objetivos de este trabajo, se trata de la filosofía del lenguaje.

A principios del XIX el filósofo danés Soren Kierkegaard afirmó que el lenguaje debería jugar un mayor papel en la filosofía occidental (Cloeren, 1998) pero fue Ferdinand de Saussure (1916) quien

sentó las bases modernas de la misma en su *Cours de linguistique générale*. Su influencia fue tal en todos los ámbitos de la filosofía y la ciencia que para describirla quedó acuñada la expresión *the linguistic turn* o vuelco lingüístico que expresa que en todas ellas se adopta un nuevo enfoque, el lingüístico, para el estudio de su objeto o en la realización de sus objetivos.

Desde este vuelco lingüístico en la documentación se vislumbran aproximaciones a conceptos tan prometedores para la clasificación e indización de contenidos como la "naturaleza del significado" de cuyo estudio Geoffrey Leech, citado por Mwihaki (2004) establece que hay dos tipos diferentes de significado, el conceptual y el asociativo. El primero de naturaleza tanto denotativa como connotativa y el segundo, asociado con la individual comprensión mental de quien habla, y que se puede subdividir en seis subtipos de contenidos: connotativos, posicionales, sociales, afectivos, reflejados y temáticos.

O la "Teoría de la referencia", que se ocupa de la interacción del lenguaje con el mundo y que según Frege citado por Picardi (2001), en la semántica contenida en cada expresión encontramos dos componentes: sentido y significado. El sentido es el pensamiento expresado, este pensamiento es abstracto, universal y objetivo. Está relacionado con los referentes y el modo en que los objetos son representados por éstos. El sentido de la frase es el pensamiento que contiene mientras que los referentes son objetos de verdad con los que el sentido se relaciona. Así la estrella de la mañana o la estrella de la tarde son dos formas de expresar, dos sentidos del mismo referente, el planeta Venus.

Así mismo, el marco que establece al considerar la estrecha relación o dependencia existente entre pensamiento-conocimiento y lenguaje se

intuye de gran interés en las tareas de clasificación documental.

Por otro lado nuevos conceptos asociados con el uso convencional del lenguaje como el "lenguaje interno" de Chomsky (1995) o la "Metasemántica" de Staiton (1996) nos abren perspectivas de estudio tan prometedoras como la del papel de la etimología o de la estilística en la formación del sentido de las palabras. O el estudio de la retórica en aspectos tan relevantes para la recuperación de información en cualquier sistema documental como son poder determinar, por ejemplo el sentido propagandístico, didáctico, peyorativo, ofensivo, etc. de los documentos a partir del uso de las palabras en ellos contenidos mediante el estudio del "dominio del discurso".⁶

Finalmente la "Teoría literaria", estrechamente ligada con la Filosofía del Lenguaje enfatiza el modo como se comprenden los textos en una perspectiva similar a la utilizada por la hermenéutica clásica (Wikipedia, 2009).

De todo lo anterior deducimos que la combinación del enfoque documental clásico basado en la filosofía de la ciencia con otro basado en la filosofía del lenguaje abre nuevas posibilidades en el tratamiento documental de la información, aportando conceptos que se intuyen valiosos para superar los obstáculos que impiden conseguir una difusión más eficaz de la información.

3.4.1 Comunicación versus contenido

Consideramos desde la perspectiva de la filosofía del lenguaje que el tratamiento documental en general y el hemerográfico en particular obtiene -hipótesis de trabajo- con el enfoque lingüístico basado en

⁶ Otro concepto que de nuevo nos aporta la filosofía del lenguaje.

la naturaleza de su función comunicativa, y que es previo al de la clasificación del conocimiento científico, una sistemática más general y más fecunda que la visión documental clásica.

Un enfoque que por sus características, las del lenguaje, puede aprehender el conjunto del flujo informativo social, ubicando en él el mensaje informativo, su génesis y evolución, y los documentos y materias que produce. Que puede más fácilmente analizar las necesidades informativas de los diferentes tipos de usuarios y por ello conectar éstos con aquéllos.

Al igual que en los demás campos informativos, en las hemerotecas, los principales problemas a resolver en la gestión eficaz de la información son el ruido y el silencio documental (Gil, 2008). El primero hace que el usuario, tras una consulta dada, tenga que enfocar su atención y dedicar un tiempo suplementario, para seleccionar, entre el conjunto de documentos devueltos por el sistema, aquellos que son pertinentes a su necesidad informativa. El segundo, más grave si cabe, consiste en la no disponibilidad de aquellos documentos que, aun siendo pertinentes no han sido devueltos por el sistema y por tanto nunca serán accedidos por el usuario.

En la literatura existente sobre el estudio de estos problemas se sitúa la polisemia del lenguaje como una de las principales fuentes de ruido, y por tanto de "infoxicación" o sobrecarga informativa.

Para evitar los efectos de la polisemia del lenguaje las clasificaciones y lenguajes documentales crean listas de términos controlados en los que se acota al máximo el valor semántico de éstos, sin embargo tanto las clasificaciones numéricas, como los tesauros o las listas de encabezamientos están lejos de ofrecer resultados

enteramente satisfactorios (Castillo, 2006). Ésto es así entre otras cosas porque hasta hace muy poco⁷ en las tareas de clasificación bibliográfica, independientemente de la herramienta clasificatoria empleada, ha sido política habitual de los centros incluir como propio de cada tema o asunto tanto los documentos que versaban sobre conocimientos propios de los temas en sí, como las posibles instancias en las que el tema se materializa. Es decir tanto los documentos que analizaban, estudiaban o trataban sobre un tema como aquellos que eran manifestaciones concretas de ese asunto. Bajo el primer prisma tendríamos la noción de **materia**, propiamente dicha, que tratan los documentos. El segundo está más relacionado con el concepto **tipo** al cual el documento pertenece o del que es un ejemplar, y que a pesar de no ser propiamente una materia, invocando el principio de uso estos documentos han sido tradicionalmente incluidos bajo el mismo encabezamiento junto con aquéllos.

3.4.2 Descriptores tipológicos y descriptores de materia

Gran parte de la difusión de la información bibliográfica (documental sería más correcto llamarla) descansa en la actualidad en los catálogos de las entidades públicas responsables de los sistemas bibliotecarios. Los OPAC de las Bibliotecas Nacionales, de las Bibliotecas Autonómicas y de las redes de lectura pública dependientes de éstas son los encargados de realizar lo que hasta los años 90 hacían las bibliografías nacionales impresas. Son ya pocas las bibliografías nacionales que se imprimen en nuestro país (BNE, 2010)⁸. Por ello las grandes distinciones por tipo de material que se realizaban en sus ediciones impresas: monografías, publicaciones seriadas, cartografía, música impresa, etc., se realizan hoy mediante subformularios del catálogo general.

⁷ Y todavía en la mayoría de los casos ya que el uso de los encabezamientos género/forma es muy reciente y en la actualidad los centros que los emplean están en una fase incipiente.

⁸ La BNE suspendió las ediciones impresas en diciembre de 2006 y en CD-ROM en septiembre de 2007

Búsqueda avanzada

Todos los campos	<input type="text"/>	Y
Autor	<input type="text"/>	Y
Título	<input type="text"/>	Y
Materia	<input type="text"/>	Y
Coleccion/Serie	<input type="text"/>	Y
Notas	<input type="text"/>	Y
ISBN, ISSN, etc.	<input type="text"/>	Y
Datos de edición	<input type="text"/>	

Buscar **Restablecer**

Idioma: CUALQUIERA

Tipo de documento: Cualquiera

Año de pub:

Ordenar por:

- Cualquiera
- Dibujos, carteles, efimera, grabados, fotografías
- Filminas, transparencias
- Grabaciones sonoras
- Kit o multimedia
- Manuscritos y archivos personales
- Mapas
- Materiales mixtos
- Monografías antiguas
- Monografías modernas**
- Partituras impresas y manuscritas
- Recursos electrónicos locales y remotos
- Prensa y revistas
- Videograbaciones

Figura 1. Delimitación de tipos documentales en el formulario de búsqueda de la BNE. Fuente: www.bne.es

Estos mismos formularios dan acceso a colecciones que hasta no hace mucho se encontraban en ficheros separados en función del tipo de material. Hoy no es raro encontrar en la misma interfaz de búsqueda los documentos de las colecciones especiales, de fondo gráfico, manuscritos, incunables y raros... etc., incluso documentos en muchos casos considerados de archivo tales como correspondencia personal, planos, memorias y proyectos de obras, etc., a toda esta variedad hay que añadir por supuesto los documentos en nuevos soportes electrónicos y digitales: CD-ROM, DVD, lápices de

memoria, discos duros portables, etc. Un conjunto de una enorme variedad tipológica cuyo contenido se separa cada día más de lo que tradicionalmente los usuarios esperaban encontrar en un catálogo de biblioteca.

Si utilizando el formulario del catálogo de la BNE buscamos p. ejemplo la expresión "guerra de las galaxias" en el título recuperamos 266 registros. De ellos solo treinta son registros de documentos en texto impreso y a primera vista no podemos saber si algunos son revistas, cuales de ellos son novelas de ciencia ficción o representaciones teatrales, cuales cómics, o si existe alguno centrado en el *making off* de la película.

Si buscamos en el mismo catálogo por el término "publicidad" en el campo de materias y acotamos por tipo de material *periódicos y revistas* nos devuelve 113 registros de los cuales no podemos saber de entrada cuales son revistas que tratan esta materia, cuales son revistas de anuncios o cuales recopilan las principales actividades desarrolladas por las agencias que trabajan en esta área.

La utilización en estos casos de los nuevos descriptores género/forma concurriendo en su caso con los tradiciones de materia podría ayudar al usuario a encontrar con mayor eficacia, seguridad y comodidad los documentos de su interés.

¿Pero de qué estamos hablando cuando hablamos de descriptores género/forma?

La primera aproximación al concepto de género en la cultura occidental lo encontramos en los tres géneros de Platón (Genette, 1979): el drama, el ditirambo y la épica que se correspondían más con la mimética e imitación en el discurso que en el contenido del mismo.

Es, sin embargo, en los estudios contemporáneos desarrollados a partir de Lloyd Bitzer (1968) donde se postula que el discurso está determinado por su situación retórica, es decir por su contexto. Según este autor es la situación la que determina la retórica del discurso y éste sólo se realiza como respuesta a una situación particular. Para él, la "situación" en relación con el discurso se compone de tres elementos: la exigencia, la audiencia y las restricciones. Elementos que de hecho establecen la relación entre el discurso y su contexto.

Carolyn Miller (1984) estudió el género desde la perspectiva de la retórica y afirmó que las *situaciones* de Bitzer eran "constructos sociales" que se materializaban en géneros. Para ella con el discurso creamos "recurrencias" o respuestas similares a través de la "construcción de tipos". En las que los tipos implican el reconocimiento de similitudes relevantes y tienen sentido sólo en relación con la interpretación de las situaciones mediante un contexto social que está fuertemente vinculado con la tradición.

Con Bawarshi (2001) se compara el género con un ecosistema. Éste determina las expectativas y el comportamiento humano dentro del mismo. Estamos modelados por el "sonido del discurso" y actuamos consecuentemente con él. Obtenemos nuestras impresiones a través del lenguaje y con él formamos nuestras acciones y percepciones.

De estas afirmaciones y otras similares, que no reproducimos por la necesaria limitación de este trabajo, podemos concluir con respecto al género las siguientes conclusiones para nuestro objeto de estudio:

- El género de los documentos nos ayuda a determinar el contexto en el que éstos actúan y por tanto su intención comunicativa.
- El contexto se modifica permanentemente al ritmo del

movimiento social, los géneros reflejan este movimiento incesante.

- La tradición, la repetición y la innovación son elementos integrantes de todo género.
- El género es reconocido antes de conocer el contenido y determina las expectativas con respecto a éste.

Podemos por todo ello y por su prometedora relación con la gestión de la información atrevernos a explorar su aplicación en el campo de la organización de la información con la vista puesta en mejorar la eficacia en su recuperación y difusión.

La aplicación de este enfoque en el área de la biblioteconomía está ya produciendo la adaptación de estos conceptos para la aplicación con los fines expuestos. Así por forma en este contexto entendemos (Sahly, 1985):

Histórica y funcionalmente la clase específica de material distinguido mediante el examen de sus características físicas, el objeto de su contenido intelectual, o la disposición de la información en su seno, por ejemplo diarios, agendas, directorios, revistas, informes, cuestionarios, planes de estudio, o calendarios.⁹

Por género nos referimos a:

Tipos específicos de materiales caracterizados por el estilo o la técnica de su contenido intelectual, por ejemplo biografías, catecismos, ensayos, himnos o comentarios.¹⁰

3.5 La recuperación de las publicaciones seriadas

La documentación objeto de las hemerotecas nacionales se delimita

⁹ La traducción es nuestra.

¹⁰ Traducción propia a partir de la definición dada en UFBD, Genre Form Field.

en la actualidad a partir del concepto de recursos continuados¹¹ que incluyen:

- Las publicaciones seriadas, que se definen como publicaciones editadas en partes sucesivas o en partes distintas, sin limitación en el tiempo y generalmente con una designación numérica [diarios, boletines, periódicos, revistas impresas o electrónicas, publicaciones anuales (informes, anuarios, repertorios, etcétera), memorias y colecciones monográficas].
- Los recursos integrados permanentes, que se definen como recursos continuados completados con actualizaciones integradas en el conjunto, sin limitación en el tiempo (bases de datos, publicaciones de hojas sueltas actualizadas de manera continua o páginas *web* actualizadas continuamente).

Las publicaciones seriadas son documentos informativamente complejos, es decir compuestos de otros documentos en los que éstos se integran: artículos, colaboraciones, noticias, ilustraciones, poemas, etc. que pueden tener entre ellos mayor o menor grado de similitud o sencillamente carecer de ella. Así en una publicación sobre novedades literarias, pongamos por caso, podemos tener desde reseñas sobre los últimos éxitos editoriales puestos a la venta, hasta algún artículo sobre la crisis en el mundo editorial o incluso algún poema o relato corto. Pero, en una revista sobre tendencias de moda, encontraremos prácticamente de todo; consejos sobre como vestir, maquillaje, dietética, o la película y la actriz de moda.

Nuestras publicaciones pueden también ser muy distintas entre sí, dependiendo de la naturaleza del fin comunicativo para el que estén concebidas. Las revistas científicas *-journals*, en el mundo anglosajón-

¹¹ La Norma ISO 3297:2007 [UNE-ISO 3297:2008] (ISSN) los define como obras puestas a disposición del público, en cualquier tipo de soporte, cuyas entregas sucesivas o integradas poseen generalmente un número de orden o una designación cronológica, y cuya duración de publicación no se ha fijado con antelación.

al estar asociadas a un objeto de conocimiento bien definido son seguramente las más fáciles de clasificar. Así *Nature* por ejemplo la encontraremos siempre cuando hagamos una búsqueda de fuentes sobre las ciencias de la naturaleza, o la *Revista General de Derecho* cuando busquemos fuentes para el Derecho. Fuera de este ámbito las dificultades para identificar materias a las que asociar publicaciones con las que podamos después recuperar fuentes informativas se complican, haciéndose a veces insalvables. Así por ejemplo, los semanarios *Tiempo* o *Interviú* nos plantearían dificultades no resueltas al menos con los sistemas de clasificación por materias generalmente usados en las bibliotecas. Igual ocurre con los centenares de boletines locales que toda biblioteca con responsabilidades territoriales atesora entre sus fondos. O las publicaciones de entidades en las que estas expresan su punto de vista respecto de multitudes de asuntos o circunstancias sean o no de actualidad. Partidos, sindicatos, asociaciones de vecinos, organizaciones empresariales, confesiones religiosas, sectas... un conjunto documental, en suma, difícil de organizar, de sistematizar para de una manera segura responder a la pregunta ¿qué publicaciones sindicales, empresariales o de partido tenemos en un momento dado en una sociedad determinada?

Encontrar un tema común que describa bien el contenido de todo un título cuando éste trata una temática heterogénea resulta difícil o poco operativo. Nos tenemos que mover entre un grado bajo de especificidad, recurriendo a términos muy generales del tipo: sociedad, cultura, etc. o bien tenemos que multiplicar los descriptores para cada publicación introduciendo un grado de dispersión que afectará multiplicando el "silencio" en recuperación.

Por otra parte nos encontramos con el problema de la colisión entre los conceptos materia, contenido sobre el que versa o trata un

documento, en este caso complejo; y género/forma, aquello que el documento es, es decir aquella categoría de la cual el documento es una ocurrencia o manifestación.

La diferenciación plena y su uso coherente contribuirán a clarificar y dar consistencia a la clasificación para una mejor difusión. Así problemas actuales de ruido y de silencio documental derivados de usar indistintamente por ejemplo términos como **cómic** tanto en estudios sobre estos documentos como para identificar este tipo de publicaciones desaparecerán.

Por ejemplo, en el caso de un usuario interesado en la historia y la evolución de las características del dibujo y su relación con los posibles contactos entre dibujantes, serán relevantes los documentos que sean ensayos que estudien el cómic, su historia, su evolución, etc. Con una búsqueda basada en el término cómic como materia generará ruido documental al proporcionar junto a éstos un conjunto de registros que no tienen que ver con su búsqueda aunque sí sean pertinentes a la misma por estar indizadas las revistas de este tipo también con ese término.

Por otra parte y en sentido contrario un usuario que quiera conocer la variedad y riqueza de esta tipología documental en un territorio y un momento determinado tendrá el mismo ruido que el anterior aunque en sentido inverso.

Sin embargo la búsqueda de soluciones a este tipo de preguntas se está ya realizando desde distintos ámbitos. De un lado, las instituciones con más reputación y autoridad en el mundo de la clasificación bibliotecaria iniciaron, ya hace años, estudios que han concluido recientemente con la propuesta de implementación de sistemas de clasificación por género/forma. De otro lado, desde el mundo de Internet, se están realizando estudios y propuestas

de clasificación de los *site web*¹² que, al menos una parte de ellos, se orientan hacia la categorización de estos recursos electrónicos atendiendo a propiedades similares -géneros *web*- a las consideradas en el mundo de la clasificación bibliotecaria.

Por todo ello y dado que la recuperación de la información en el ámbito concreto de las publicaciones seriadas se enfrenta con dificultades similares a las de la recuperación de los recursos informativos en Internet nuestra propuesta consiste en aplicar lo mejor de los dos mundos: la clasificación bibliotecaria y los avances en la recuperación *web* basados en la utilización de recursos de inteligencia artificial para la identificación y asignación automática de esta clasificación.

¹² Documentos con estructura temática compleja y cambiante, en muchos aspectos, similar a la de las publicaciones seriadas.

4. Realización de la prueba

4.1 Análisis y descripción de modelos

En el contexto de este trabajo, llamamos modelos orientados hacia el contenido a aquellos como el de la BNE¹³ en los que la recuperación por materias de sus documentos se basa en la utilización, durante la indización y para las búsquedas, de términos (encabezamientos de materia) que pueden ser utilizados indistintamente para describir la materia de los documentos y su tipo.

Por el contrario llamamos modelos orientados a la comunicación, también en este contexto, a aquéllos que basan la recuperación de los documentos en su papel en la comunicación social¹⁴. Consecuentemente con eso implementan estrategias para reducir la ambigüedad de aquellos términos que son susceptibles de transmitir el doble significado mencionado. Para ello separan los encabezamientos en dos ficheros o listados independientes, el de materias que contiene términos utilizables exclusivamente para documentos que “tratan sobre” la materia en cuestión, y el de género/forma que expresa aquello que el documento “es”, (en la comunicación social) independientemente de cual sea su materia. Así mismo estos sistemas codifican también separadamente unos y otros encabezamientos mediante el uso de campos MARC separados. El 650 (materias) y 655 (género/forma) de modo que la recuperación pueda realizarse con absoluta precisión.

¹³ En el momento de la redacción de este trabajo (marzo 2011) la BNE ha empezado a utilizar los encabezamientos género/forma, si bien aún no es posible realizar búsquedas a partir de los mismos.

¹⁴ En el sentido que se expresa en la introducción, pág. <6>. Es decir en la práctica social mediante la cual los grupos humanos establecen el contexto, las condiciones y los objetivos de su interrelación. Abarca todas las actividades humanas.

El modelo clásico utilizado para la difusión de la información bibliográfica es el utilizado hasta la fecha en las bibliografías nacionales, es decir aquel que se basa en la clasificación de los registros en base a sus materias y/o números CDU.

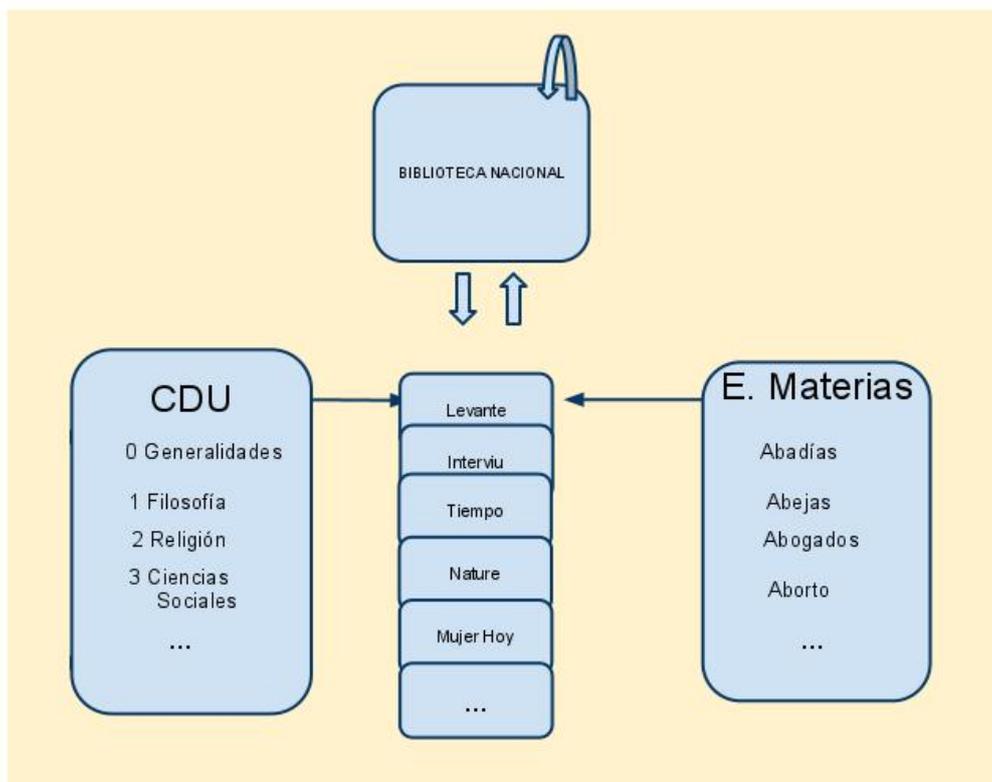


Figura 2. Modelo A o clásico. Fuente: elaboración propia

Los problemas de este enfoque en el caso de las publicaciones seriadas están relacionados con la naturaleza compleja de éstas desde el punto de vista de las materias que tratan. De hecho, este modelo se pensó para la clasificación y recuperación en las bibliotecas de los documentos que han sido desde siempre su tipología principal, los libros. Es decir para obras unitarias, acabadas y monográficas desde el punto de vista de su materia. Después se extendió a otros tipos de documentos en los que como las publicaciones seriadas, su funcionamiento es ineficiente. Por ello en las instrucciones de uso de estos mismos lenguajes, suele advertirse que su uso no es eficaz en obras de creación literaria y en general en aquellas que carecen de materia, así como en las publicaciones seriadas por ser documentos

que cuando las tienen éstas suelen ser múltiples.

Exceptuando las revistas científicas especializadas, tipología con un peso porcentual minoritario en el conjunto de estas publicaciones, la mayoría de las mismas, se encuentran en esta situación. Bien tratan de materias múltiples y variadas en razón de los diversos artículos que las componen, o no tratan de ninguna, por ser su objeto otro, como la publicidad de productos y servicios, o el entretenimiento y la distracción.

Por ello, y porque a éstos se suman además los problemas derivados del ruido documental que producen, se hacen necesarias soluciones que mejoren la respuesta ofrecida a los usuarios por los sistemas de recuperación de documentos.

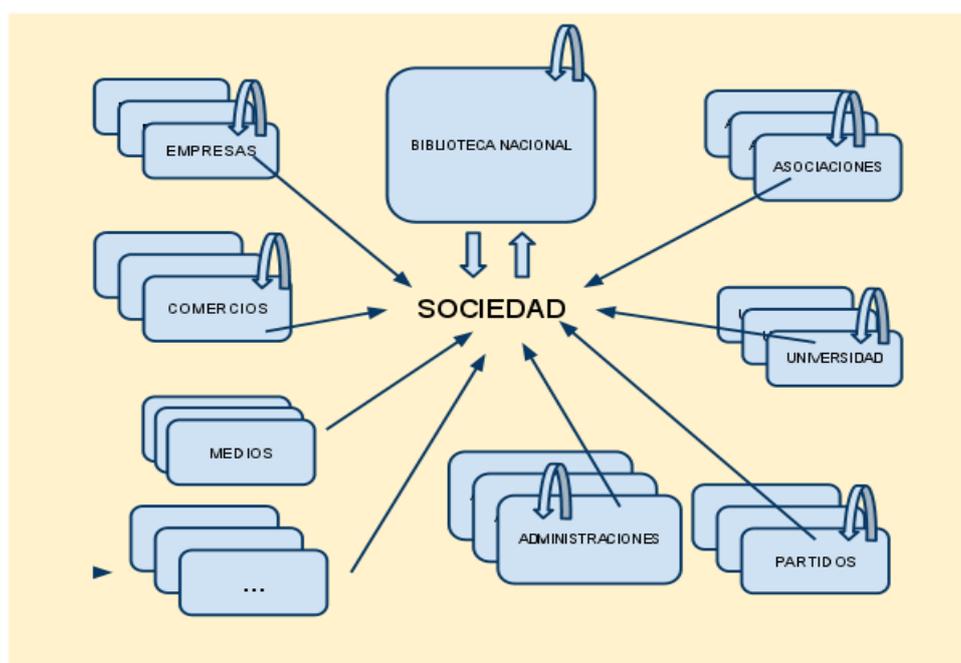


Figura 3. Entidades Modelo B. Fuente: elaboración propia

Se trata de un modelo que con respecto a los lenguajes documentales clásicos ha sido enriquecido semánticamente mediante la

adopción de un nuevo enfoque que pueda facilitar el trabajo de los ordenadores en la interpretación del sentido informativo de los documentos con la identificación automatizada de su género y forma.

En este nuevo enfoque, se desplaza el centro del interés clasificatorio desde el contenido de los documentos (su materia) hasta otra categoría, el género-forma, que por estar imbricada en el centro de la perspectiva de la comunicación social, permitirá, esa es la hipótesis, mejorar la capacidad descriptiva, y con ella los resultados, del modelo de clasificación/recuperación aplicado a los ítem informativos, en nuestro caso los registros de los documentos propios de las hemerotecas.

Además, la definición de una ontología semántica con categorías de las que se desprendan los descriptores que integran el lenguaje pueda aspirar, en su desarrollo a establecer las reglas de inferencia necesarias para, por una parte asignar automáticamente los descriptores a los documentos, y por otra, actualizar dinámicamente las categorías y sus descriptores en sintonía con los cambios culturales, sociales y lingüísticos que se vayan produciendo.

En la ontología del modelo clasificatorio expuesto, existen entidades tales como empresas, comercios, administraciones, etc., que se relacionan consigo mismo, con las de su clase o con toda la sociedad, utilizando diferentes medios en función de sus objetivos de comunicación. Así, en sus comunicaciones internas, las entidades y empresas generan: boletines, diarios de empresa, intranets o redes internas... etc., ésto es, documentos por los que fluyen información y noticias relacionadas con el interés de la entidad o con su organización. Debido a que estos documentos suelen tratar sobre temas múltiples y diversos; y a que es su enfoque lo que los agrupa, clasificarlos por su(s) materia(s) además de no ser muchas

veces posible, tiene menos sentido que hacerlo por la función que desempeñan.

En su relación con las entidades de su clase generan *newsletter*, boletines, revistas o diarios especializados en el campo de actividad de la entidad. Pueden tener *web* con o sin extranet. Por las mismas razones que el caso anterior no se deben clasificar por materias.

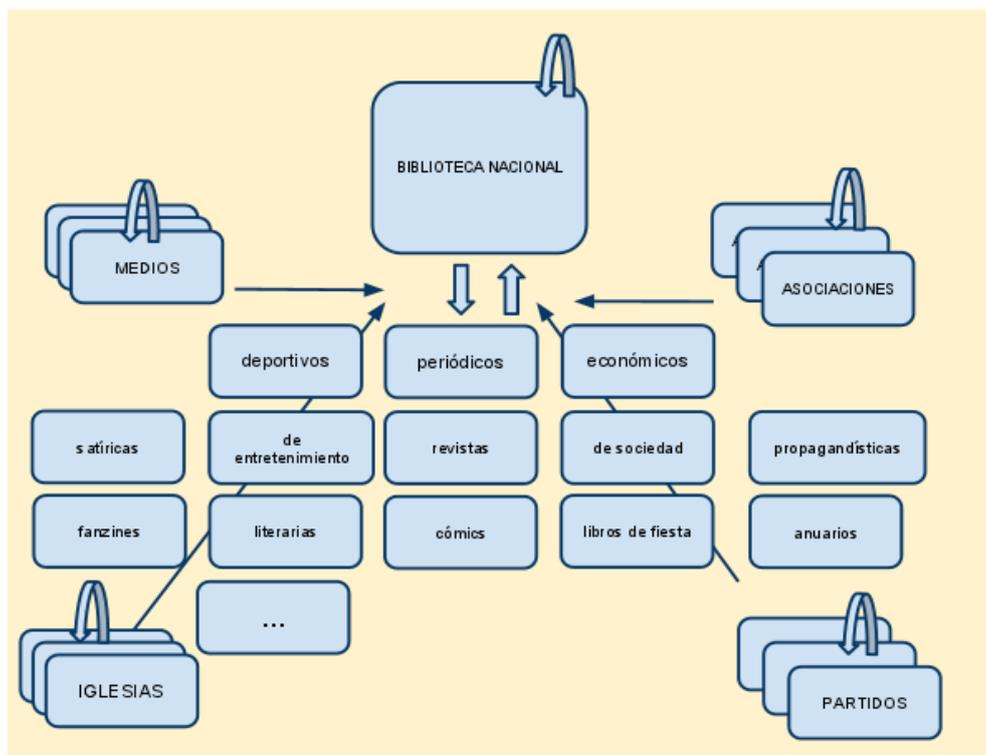


Figura 4. Tipos documentales del modelo B. Fuente: elaboración propia

Finalmente en su relación con la sociedad, generan revistas de divulgación general y pueden desarrollar webs o portales. En algunos casos se pueden clasificar por materias y en todos, por el género al que pertenecen: revistas especializadas, de divulgación, de opinión...

En las entidades del tipo *empresa* y sobre todo en las de tipo *comercios*, como consecuencia de su función principal, pueden desarrollar un medio característico que es el catálogo de productos o servicios que ofrecen. Éste, cuando lo realizan entidades terceras o intermediarios, pueden llegar a ofrecer diarios, revistas, webs o

portales con las ofertas de productos o servicios de todo un sector de la economía, por ejemplo revistas de compra venta de inmuebles, de vehículos, entre particulares (*El Trajín, El Baúl...*). Éstas son publicaciones, obviamente sin materia. En realidad, una forma desarrollada de catálogo de productos o servicios.

Otras entidades del tipo empresa, del tipo medio de comunicación, del tipo asociación o de un tipo que es combinación de dos o más de ellas, pueden ofrecer también productos o servicios, en este caso informativos o culturales como: noticias de actualidad; o de análisis y opinión en profundidad sobre actualidad; o sobre actividades realizadas o venideras: memorias, anuarios, etc. las primeras; programas de fiestas: patronales, de fallas, de moros y cristianos, etc., de certámenes, etc. las segundas.

Otras comunican creaciones artísticas originales, como poemas, o relatos cortos. Otras ofrecen relatos gráficos como los cómics con fines de entretenimiento, de humor, incluso con fines artísticos en si mismos. Otras sirven para difundir un planteamiento o un punto de vista político, o filosófico o religioso, o para criticar a veces con sátira, escarnio o burla el del contrario.

4.2 Análisis de descriptores según su ambigüedad

Son muchos los términos y expresiones utilizadas en los lenguajes documentales clásicos susceptibles de albergar múltiples significados. Tanto es así que en las instrucciones de uso de muchos lenguajes de clasificación se dan normas concretas sobre cómo tratar con este problema. Normas que como veremos resultan insuficientes para las necesidades informativas actuales. Vamos a ver algunos casos extraídos del fichero de autoridades de la BNE y a analizar su comportamiento en una búsqueda bibliográfica.

En el fichero de autoridades de la BNE, pero también en otros de los principales ficheros y listas de autoridades de materia usadas en España: LEMAC, LEMBP., fichero de autoridades del CSIC... etc. encontramos con las siguientes categorías de términos:

- Términos, o combinaciones de términos, que expresan una materia, y que su utilización está expresamente restringida para describir documentos que tratan *sobre* la misma. En principio estos términos deberían ofrecer resultados pertinentes en las búsquedas en un porcentaje alto. No han sido estudiados, en este proyecto, más allá de la constatación de su existencia.

Cambiar Formato ▼

Formato: Etiquetado

Acústica

Úsase únicamente para los manuales teóricos de esta ciencia

Término

genérico: Física

Término

específico: Acústica arquitectónica
Acústica musical
Acústica submarina

Término

relacionado: Ondas sonoras
Sonido
Ultrasonidos
Vibraciones
Electroacústica
Ruido

Fuentes: LCSH (Sound)
LEM

Nº Registro: XX524570

Figura 5. Encabezamientos tipo 1. Fuente: Catálogo Autoridades BNE

- Términos, o combinaciones de términos, que pueden expresar tanto una materia como un tipo documental, y que, bien expresa, bien implícitamente, están autorizados para ello en el sistema. Estos son los términos que más nos interesan por ser sospechosos de ser fuente de ruido en las búsquedas. Se ha estudiado su comportamiento mediante un análisis estadístico exhaustivo.

Almanaques

Úsase para publicaciones anuales que recogen datos, noticias o escritos de diverso carácter, así como para los estudios sobre los propios almanaques

Término

genérico: Publicaciones menores
Publicaciones seriadas

Término

relacionado: Calendarios
Relaciones de sucesos

Fuentes: LCSH (Almanacs)
RAMEAU (Almanachs)

Nº Registro: XX524985

Figura 6. Encabezamientos tipo 2. Fuente: Catálogo Autoridades BNE

- Términos cuyo uso está autorizado en exclusiva en documentos que "son" colecciones, antologías o concursos. También pueden ser fuente de ruido, aunque en este caso por desconocimiento o incoherencia en la aplicación de sus restricciones. Hemos estudiado una muestra de ellos.

Autobiografías

Úsase para colecciones, antologías, concursos, etc. Para las obras sobre este género literario, véase: Autobiografía

Término

genérico: Biografías

Término

relacionado: Autobiografía

Fuentes: LCSH (Autobiographies)

Nº Registro: XX525482

Figura 7. Encabezamientos tipo 3. Fuente: Catálogo Autoridades BNE

- Términos que están expresamente indicados para describir los documentos desde el punto de vista de su tipo, no permitiéndose su uso para la indización de materias. Su cantidad es claramente insuficiente para las necesidades informativas actuales. A causa de ello, y sobre todo, a que su uso no está prescrito en un campo específico y separado, el de género/forma (MARC 655) las búsquedas que pretenden satisfacer necesidades informativas del tipo: cuantas revistas de anuncios entre particulares, o cuantas carteleras de espectáculos existen en el fondo documental, o cuantos cómics, o cuantas revistas satíricas valencianas del siglo XIX... etc., son imposibles de satisfacer. Estamos pues, en relación con estos encabezamientos también ante un problema de silencio documental. Problema que por los límites de este proyecto la evaluación de su alcance queda fuera del mismo.

Bibliografías

Úsase para obras que consisten en listas de documentos publicados. Para obras sobre la bibliografía como ciencia, véase: Bibliografía

Véase además el subencabezamiento: Bibliografías bajo determinadas materias, ej.:
Medicina-Bibliografías, Cervantes Saavedra, Miguel de-Bibliografías, España-Bibliografías

Término

genérico: Obras de referencia

Término

específico: Bibliografías críticas
Bibliografías de bibliografías
Bibliografías nacionales
Bibliografías universales
Biobibliografías

Término

relacionado: Catálogos de editores
Catálogos de libreros
Libros seleccionados
Bibliografía
Bibliotecas
Índices

Fuentes: UNESCO (Bibliografías)

Nº Registro: XX528151

Figura 8. Encabezamientos tipo 4. Fuente: Catálogo Autoridades BNE

Los problemas de polisemia-ruido (y también de silencio) son inabordables con este tipo de sistemas por ser de una parte imprecisos (hay términos para los que se permite tanto materias como tipos) y de otro lado demasiado complejos para ser conocidos y empleados correctamente por un usuario medio; y que en cualquier caso resultan ineficientes como sistema de recuperación de información para los grandes y complejos repositorios de documentación actuales.

Tipos de encabezamientos	
Tipo 1	acústica, antropología física, archivística, arqueología, armas, autobiografía, auto sacramental, bacteriología, bibliografía, biografía, bromatología, botánica, catálogos de bibliotecas, cuento, comedia, empresas, ensayo, entremés, epigrama, novela, ópera, poesía, religión, sainete, sátira, sectas, tragedia
Tipo 2	almanaques, asociaciones, baladas, calendarios, cartas, cartas de amor, catálogos de editores, catálogos comerciales, catálogos de librerías, catálogos de subastas, catecismos, cómics, guiones radiofónicos, guiones de televisión, guiones cinematográficos, chistes, efímeros, flamenco, folclore, jarchas, juegos de mesas, libros litúrgicos, música para piano, publicaciones periódicas, presupuestos, tesauros,
Tipo 3	autobiografías, autos sacramentales, biografías, comedias, cuentos, ensayos, entremeses, , epigramas, normas técnicas, novelas, películas cinematográficas, poesías, sainetes, sátiras, tragedias
Tipo 4	bibliografías, -bibliografías, -biografías, -catálogos, -cómics, -entrevistas, -guiones radiofónicos, -guiones de televisión, -guiones cinematográficos, -libros litúrgicos, -normas, -novelas, zarzuelas, óperas

Tabla 1. Tipos de encabezamientos. Fuente: elaboración propia

Los términos del primer tipo pueden ser relativamente eficaces para la recuperación de documentos que *tienen materia* siempre que ésta no sea compleja y que su denominación no sea al mismo tiempo la de un tipo documental habitual en el contexto en el que se da la búsqueda/recuperación, son términos del tipo: arquitectura, agua, frutas, incendios...

Los del segundo tipo generan inevitablemente ruido en las respuestas del sistema al aparecer en ellas indistintamente documentos sobre una materia con documentos que no lo son pero que han sido descritos con el mismo término que aquélla. Más adelante veremos ejemplos de estos términos y cuales son los resultados cuando aparecen involucrados en las búsquedas.

Los términos del tercer tipo, podrían ser útiles para separar unos documentos (con materia) de otros (instancias de un determinado tipo) pero al limitarse su uso para agrupaciones de documentos se pierde esta posibilidad. El resultado es que no podemos saber cuantos documentos tenemos en un determinado fondo o colección

de cuentos por ejemplo, ya que este encabezamiento tiene limitado su uso a colecciones, antologías, concursos, etc., ni de propiedad inmobiliaria-compraventa, puesto que esta combinación de términos puede dar resultados válidos tanto para publicaciones que tratan sobre estudios jurídicos o económicos sobre la compraventa de viviendas como para aquellos periódicos especializados en poner en contacto la oferta con la demanda de estos bienes.

Finalmente los términos del cuarto tipo, son a nuestro entender los que se deberían desarrollar suficientemente hasta cubrir toda la casuística tipológica del fondo. Además, constituir con ellos un fichero de autoridad, separado del fichero de materias, para poder de este modo recuperar los documentos a partir de sus atributos tipológicos, aún cuando éstos, se denominen o no de igual modo que su materia (polisemia). Y/o cuando carezcan de la misma.

4.2 Determinación de las variables de comparación de la prueba

Hemos analizado el comportamiento de un conjunto de encabezamientos de materia seleccionados¹⁵, para a partir de los resultados obtenidos determinar el índice de ruido que producen.

Considerando como hipótesis del trabajo que el uso de estos encabezamientos en las búsquedas (variable independiente de la hipótesis) produce un índice de ruido (variable dependiente) mucho mayor que con la aplicación de los encabezamientos género/forma adecuados en un sistema de recuperación que los implemente.

Para comprobar nuestra hipótesis, hemos comparado los resultados

¹⁵En el ámbito de la recuperación de registros bibliográficos en general y de los recursos continuados (publicaciones periódicas y seriadas) en particular.

obtenidos a partir de las consultas en un sistema clásico como es el de la BNE, con los obtenidos en otro sistema de recuperación, el de la hemeroteca de la Biblioteca Valenciana, en el que éstos se están empleando, y, aunque en una fase experimental, son posibles ya las búsquedas por ellos.

Dado que esta aplicación es muy reciente y el número de registros disponibles es reducido, el valor de esta comparación quedará pendiente de ser comprobado de manera completa en trabajos futuros.

Por otra parte hemos seleccionado del fichero de autoridades de materia de la BNE (2011) un conjunto de encabezamientos del segundo tipo que han sido utilizados para seleccionar registros de este catálogo. El objetivo era comprobar para cada caso el ruido generado. Las búsquedas han tenido en común el objetivo expreso de obtener, por una parte los documentos que tratan sobre la materia buscada, son los pertinentes en esta búsqueda, y por otra aquellos que son instancia de un tipo documental determinado, asociado o no a esa materia.

La metodología utilizada ha consistido en considerar para cada uno de los términos escogidos si los documentos devueltos por el sistema se correspondían o no con su materia, es decir, si "trataban sobre" ella o si por el contrario "eran instancias" es decir un ejemplar de ese tipo de documento. En el primer caso se han considerado pertinentes para la búsqueda y en el segundo, no. De ahí que las tasas de pertinencia y ruido sean complementarias y se inviertan según la consulta pretenda obtener materias o tipos.

Cuando los documentos devueltos por la búsqueda excedían del centenar, para la realización de los diferentes análisis se han obtenido

muestras representativas que permiten obtener los resultados con un margen de error de ± 9 puntos porcentuales, trabajando con un nivel de confianza del 95 %.

4.4 Resultados en el catálogo de la BNE

Encabezamiento	n. de docs.	muestra	pertinentes	tasa pert.	tasa ruido
Almanaques	10	10	1	0,1	0,9
Anarquismo	39	39	3	0,08	0,92
Calendarios	274	83	0	0	1
Catálogos de subastas	39	39	0	0	1
Creación literaria	9	9	4	0,44	0,56
Cómic	360	89	11	0,12	0,88
Humorismo	44	44	0	0	1
Literatura española	345	88	24	0,27	0,73
Partidos políticos	22	22	1	0,27	0,73
Poesía / poesías	341	88	8	0,09	0,91
Propiedad inmobiliaria	15	15	1	0,07	0,93
Psicoanálisis	55	55	53	0,96	0,04
Publicaciones periódicas	226	78	51	0,65	0,35
Presupuestos	357	89	19	0,21	0,79
Partidos políticos	23	23	10	0,43	0,57
Publicidad	113	59	49	0,83	0,17
Sindicatos	134	63	7	0,11	0,89
			Tasa Media	0,27	0,73

Tabla 2. Comportamiento de términos del modelo A. Fuente: elaboración propia

Veamos estos resultados caso por caso. En la primera búsqueda con el requerimiento almanaques en el campo materias; y, prensa y revistas como tipo de documento,

Búsqueda avanzada

Todos los campos Y

Autor Y

Título Y

Materia Y

Coleccion/Serie Y

Notas Y

ISBN, ISSN, etc. Y

Datos de edición

Buscar **Restablecer**

Idioma:

Tipo de documento

Año de pub:

Ordenar por:

Figura 9. Formulario de búsqueda en BNE. Fuente: www.bne.es

obtenemos un resultado de diez registros.

Resultados de la búsqueda

Materia "almanaques" la búsqueda recuperó 10 Registros bibliográficos.

#1	PENDIENTE DE SIGNATURA	195
<input type="button" value="Detalles"/>	Al'manah "Gomonu Ukraïni" [Texto impreso]	
<input type="button" value="Guardar"/>	1 ejemplar disponible en Sede de Recoletos en Sala de Prensa y Revistas-Petición anticipada	
#2	AHS/35921	1902
<input type="button" value="Detalles"/>	Almanaque-guía de Guadalajara [Texto impreso] : centros oficiales y particulares, agricultura, comercio, industria, profesiones, anuncios	
<input type="button" value="Guardar"/>	2 ejemplares disponibles en Sede de Alcalá	

.....

Figura 10. Documentos devueltos por la búsqueda. Fuente: www.bne.es

Cuando los analizamos uno por uno encontramos que nueve de los diez documentos devueltos se corresponden con registros que son instancia del concepto almanaques y sólo uno de ellos se puede considerar como una obra que trata sobre almanaques, de hecho hace una compilación, entendemos que anotada, de este tipo de documentos.

#8	HNB/11161	186
Detalles	Almanaque político y literario de La Soberanía nacional [Texto impreso]	
Guardar	1 ejemplar disponible en Sede de Recoletos en Sala de Prensa y Revistas	
#9	HNB/11043	186
Detalles	Colección de almanaques para ... [Texto impreso] : compuestos por una sociedad de literatos	
Guardar	1 ejemplar disponible en Sede de Recoletos en Sala de Prensa y Revistas	
#10	ZR/	18
Detalles	Almanaque diabólico [Texto impreso]	
Guardar	1 ejemplar disponible en Sede de Alcalá en Ejemplar de conservación	

Figura 11. Documento pertinente. Fuente: www.bne.es

4.5 Resultados en el catálogo de la BV

Paso 1: Búsqueda por encabezamiento almanaques, en campo, término género/forma¹⁶

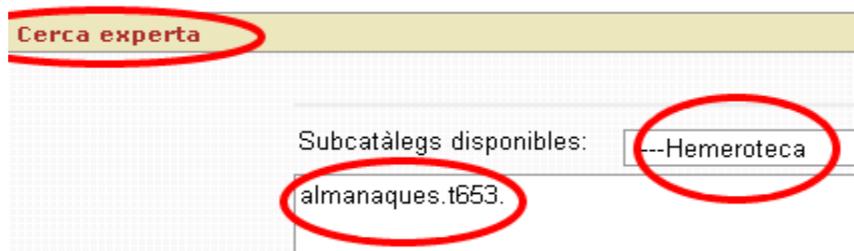


Figura 12. Búsqueda por tipo documental en hemeroteca BV. Fuente: bv.gva.es

Resultado:

Hemeroteca - Registres trobats: 25 (1/25)	
<input type="checkbox"/>	1 Almanac de La Traca (1914)
<input type="checkbox"/>	2 Almanac... l'esquella de la torratxa (1924)
<input type="checkbox"/>	3 Almanak mercantil o guía de comerciantes para el año de... por D.D.M.G. (1799)
<input type="checkbox"/>	4 Almanaque agenda de La Agricultura española para el año ... (1901)
<input type="checkbox"/>	5 Almanaque cultural... / del Circulo de Lectores (1990)
<input type="checkbox"/>	6 Almanaque de la ilustración para el año de... (1873)
<input type="checkbox"/>	7 Almanaque de El Liberal de Valencia para... (1902)
<input type="checkbox"/>	8 Almanaque de El Clarín (1923)
<input type="checkbox"/>	9 Almanaque de las artes y las letras para... (1928)
<input type="checkbox"/>	10 Almanaque de Corcel para... (1940)
<input type="checkbox"/>	11 Almanaque del bello sexo : para el año bisiesto de... (1864)
<input type="checkbox"/>	12 Almanaque Destino... (1944)
<input type="checkbox"/>	13 Almanaque La Vanguardia. (1960)
<input type="checkbox"/>	14 Almanaque Martí [Texto impreso] (1914)
<input type="checkbox"/>	15 Almanaque para todo el mundo (1889)
<input type="checkbox"/>	16 Almanaque-agenda agricola : 1905 (1905)
<input type="checkbox"/>	17 Anuario Soler... : Almanaque-Agenda de bolsillo [y] Guía de Castellón y de toda la provincia (1901)
<input type="checkbox"/>	18 Calendari ;Cu-cut! pera ... (1902)
<input type="checkbox"/>	19 Calendario para el Reyno de Valencia... almanaque ilustrado (1857)
<input type="checkbox"/>	20 Calendario para... El Mercantil Valenciano a sus suscriptores (1882)
<input type="checkbox"/>	21 Pumby : almanaque (1958)
<input type="checkbox"/>	22 Valencia cómica : almanaque año... (1891)
<input type="checkbox"/>	23 La Lauset : armanac dal patriot lengodoucian... per l'an ... / co-lauraires Jean Aicard ... [et al.] (1877)
<input type="checkbox"/>	24 Lo Xanguet : o sia almanach per l'any ... (1864)
<input type="checkbox"/>	25 Der Sternkalender auf das Jahr ... (1949)

Figura 13. Documentos devueltos en la búsqueda de la fig. 12. Fuente: bv.gva.es

¹⁶Estos términos están indizados provisionalmente en el campo marc 653 mientras se implementa el campo marc 655 definitivo.

25 registros.
Todos *son* almanagues.
Tasa de pertinencia 100 %.
Tasa de ruido 0%

Paso 2. Búsqueda por encabezamiento almanagues en campo materias

The screenshot shows a search interface with a header 'Cerca experta'. Below it, there is a section for 'Subcatàlegs disponibles:' with a dropdown menu set to '----Hemeroteca'. A search input field contains the text 'almanagues.t650.'. Red circles highlight the header, the dropdown menu, and the search input field.

Figura 14. Búsqueda por materia en hemeroteca BV. Fuente: bv.gva.es

Resultado:

The screenshot shows the search results page. The header 'Cerca experta' is followed by 'Resultats'. A red circle highlights 'Resultats' and an arrow points to a message box that says 'No hi ha cap registre amb aquestes característiques'. Another red circle highlights this message box.

Figura 15. Documentos devueltos en búsqueda de la fig. 14. Fuente: elaboración propia (captura del formulario del OPAC de la BV)

0 registros.
No hay registros de publicaciones *sobre* almanagues
Tasa de pertinencia 100 %.
Tasa de ruido 0%

4.6 Resultados de la prueba

En el caso de la BNE, con los encabezamientos seleccionados la tasa de pertinencia de los resultados cuando se buscan documentos *sobre esa materia* oscilan entre el 0 de los calendarios o los catálogos de subastas y el 0,96 de las revistas de psicoanálisis, situándose la media de pertinencia en 0,27.

Si por el contrario consideramos que nos interesan documentos que son instancia del encabezamiento buscado los resultados obviamente son los mismos aunque invertidos, es decir 1 de los calendarios, catálogos de subastas o revistas de humorismo, y el 0,04 de psicoanálisis, con 0,73 de media global.

En el caso de la hemeroteca de la BV, para el caso concreto del término almanaques, seleccionado expresamente para compararlo con el mismo término en la BNE, los resultados son para el tipo documental: 25 ítem pertinentes de 25 registros (100%); y para el tópico almanaques: 0 ítem pertinentes de 0 registros (100%).

Los resultados obtenidos del análisis de las muestras y de la comparación de los dos sistemas de recuperación confirman que los sistemas clásicos de recuperación de documentos son, para este tipo de búsquedas, muy ineficientes. La tasa de pertinencia de 0,27, es en sí mismos, elocuente. Si además la comparamos con un sistema que permite la recuperación separada de los conceptos materia y tipo, en el que como hemos visto los resultados son 100% de pertinencia con 0% de ruido, las conclusiones son, a nuestro entender, inapelables.

Por otra parte, el hecho de que nuestras búsquedas estén limitadas, a causa del alcance de este trabajo, a las publicaciones seriadas, pone en marcha inadvertidamente una restricción contextual que podemos

observar y analizar cuando cambiamos la búsqueda por ejemplo, en monografías modernas.

The image shows a list of 10 search results for almanacs. Each entry includes a call number, a 'Detalles' button, the title and author, a 'Guardar' button, and the location/availability. Entry #5, 'Almanaque de la ilustración para el año de ...', is highlighted with a red rectangular box. A 'Resumen' link is visible at the bottom right of the list.

#	Call Number	Title	Author	Availability
#1	BM 050.9(450)	Almanacchi, lunari e calendari toscani tra Settecento e Ottocento [Texto impreso] 1ª ed.	Solari, Gabriella	1 ejemplar disponible en Sede de Recoletos en Sala Información Bibliográfica
#2	B 77 ALE COBURG	Die Almanache, Kalender und Taschenbücher (1750-1860) der Landesbibliothek Coburg [Texto impreso]	Landesbibliothek Coburg	2 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá
#3	VC/12361/10	Les almanachs populaires aux XVII>2< et XVIII>2< siècles [Texto impreso] : Essai d'histoire sociale	Bollème, Geneviève	1 ejemplar disponible en Sede de Recoletos en Salón General
#4	SDB 82-91 BOL	Les almanachs populaires aux XVIIe et XVIIIe siecles [Texto impreso] : essai d'histoire sociale	Bollème, Geneviève	2 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá
#5	12/713725	Almanaque de la ilustración para el año de ... [Texto impreso]		1 ejemplar disponible en Sede de Alcalá en Salón General-Petición anticipada
#6	VC/17277/3	Almanaque [Texto impreso] : tempos e saberes	Radich, Maria Carlos	1 ejemplar disponible en Sede de Recoletos en Salón General
#7	BA/17535	Les anciens almanachs illustrés [Texto impreso] : Histoire du calendrier depuis les temps anciens jusqu'a nos jours	Champier, Victor	2 ejemplares disponibles en Sede de Recoletos
#8	10/116877	Los antecedentes del periódico de masas: ocasionales, "canards" y almanaques [Texto impreso] 1ª ed.	Núñez de Prado y Clavell, Sara	3 ejemplares disponibles en Sede de Alcalá
#9	B 44 CHEC	Bibliografia slovenskych a inorecovych kalendárov [Texto impreso] : 1701-1965	Kipsová, Mária	1 ejemplar disponible en Sede de Recoletos en Sala Información Bibliográfica
#10	E/3603	Calendaris i almanacs [Texto impreso]	Vélez, Pilar 1957-	2 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá

Figura 16. Resultados devueltos para almanaques en monografías. Fuente: www.bne.es

En esta selección de resultados podemos ver que si bien siguen

apareciendo mezclados tanto almanaques como obras sobre los mismos, éstas son de una variedad mayor puesto que encontramos tanto recopilaciones como estudios históricos o ensayos sobre los mismos. La explicación de esta circunstancia puede estar en que el tipo documental monografías sea más adecuado para tratar sobre una materia, en este caso, almanaques, que el tipo documental revistas. Del mismo modo que para tratar por ejemplo sobre cine, el tipo documental **monografías** sea también más fecundo que el tipo **películas**, aunque pueda haber películas que tengan como materia una película o cualquier otro asunto.

Por otra parte, aunque en los sistemas clásicos evaluados, éstos se comportan del mismo modo en todas las búsquedas, es decir, sin diferenciar los documentos sobre determinadas materias de los que son en realidad instancias de la misma, hay no obstante, situaciones y casos diversos que será conveniente analizar en trabajos de mayor alcance. Así, por ejemplo en términos como anarquismo o psicoanálisis a diferencia de almanaques, la polisemia de los mismos resulta menos evidente. En ellos lo que hay que dilucidar es si ante nosotros tenemos el discurso psicoanalista o anarquista, respectivamente, o si bien se trata del metadiscurso sobre éstos. Solo en el segundo caso se deberían considerar materia y para el primero necesitaremos de encabezamientos género/forma perfectamente separados para evitar el ruido en los resultados de las búsquedas.

#6	HN/2026 (MICROFILMADO)	1885
Detalles	Bandera social [Texto impreso] : semanario anárquico-colectivista	
Guardar	3 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá	
#7	HNA/3828	1993
Detalles	Bicel [Texto impreso] : boletín interno del Centro de Estudios Libertarios Anselmo Lorenzo	
Guardar	Bicel. Boletín interno del Centro de Estudios Libertarios	
	2 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá	
#8	ZA/1698	1988
Detalles	Boletín bibliográfico [Texto impreso]	
Guardar	Boletín bibliográfico (Ateneu Enciclopèdic Popular-Centre de Documentació Històrico-Social)	
	1 ejemplar disponible en Sede de Alcalá en Sala de Prensa y Revistas-Petición anticipada	
#9	D/22234	1999
Detalles	Butlletí bibliogràfic [Texto impreso]	
Guardar	1 ejemplar disponible en Sede de Recoletos en Sala de Prensa y Revistas	
#10	D/9223	1983
Detalles	Cénit [Texto impreso] : órgano de la C.N.T.-A.I.T. Regional del Exterior, portavoz de la C.N.T. de España	
Guardar	Cénit (Toulouse. Semanal)	
	2 ejemplares disponibles en Sede de Recoletos y Sede de Alcalá	

Figura 17. Resultados para búsqueda 'anarquismo' en subformulario: 'prensa y revistas'. Fuente: www.bne.es

De los 39 resultados obtenidos para esta materia, solo tres son un metadiscurso anarquista, es decir, documentos que tratan sobre el discurso anarquista. Si bien en este caso son además y al mismo tiempo un tipo documental concreto: una bibliografía.

Vemos pues que las relaciones entre materia y tipo (instancia de) y, materia y discurso (instancia de) no son en absoluto estáticas ni uniformes. Tenemos en los dos casos vistos unos documentos (compilación, y bibliografías) que siendo en sí mismos más formales que sustantivos pueden actuar y de hecho actúan como materia de

otros.

Es el contexto de la comunicación o el contexto *en la* comunicación el que determina la función de objeto o sujeto de los documentos. Función que importa y mucho a los usuarios a la hora de encontrar los documentos que necesitan para solucionar una falta de información o una necesidad de conocimiento específico.

Conseguir sistemas de recuperación de información que avancen en la determinación del contexto y con ello mejoren los resultados actuales se evidencia como un reto no solo para las bibliotecas sino también y principalmente para aquellos sistemas que como los motores de búsqueda en Internet adolecen de las mismas o de mayores deficiencias.

La clave está en introducir mecanismos de contextualización del diálogo entre el usuario y los fondos gestionados por el sistema. Uno de de estos mecanismos puede ser la determinación del contexto de la necesidad informativa. Qué tipo de información, contenida en qué tipo de documentos puede satisfacerla.

Las monografías, seguramente junto con los artículos de revista, sean los tipos documentales que mayoritariamente comuniquen todo tipo de materias. Ahora bien ni todas ellas tratan siempre exclusivamente de materias, lo vemos en la ilustración (figura 16) en la que para la materia almanaques nos devuelve el ítem 5 que es un almanaque, ni son los únicos tipos documentales que los tratan, puesto que podía darse el caso de por ejemplo películas, vídeos o registros sonoros *sobre* los mismos.

De todo ello podemos concluir que para mejorar el resultado de las búsquedas en los actuales catálogos integrados por múltiples

tipos documentales, la identificación de los roles materias-tipologías que estos desempeñan en cada caso podría aportar un contexto de desambiguación de términos con fuertes tendencias polisémicas mejorando por ello la pertinencia de los resultados al reducir el ruido en la búsqueda.

5. Conclusiones

Hemos descrito a partir de la formulación conceptual de dos modelos teóricos de tratamiento y recuperación de documentos en general, y de publicaciones seriadas en particular, las carencias del modelo clásico. Se ha esbozado la posibilidad de superar estas carencias mediante la implementación de un modelo más orientado a contemplar la naturaleza y el tipo de los documentos, muy determinados por la función concreta que desempeñan en el flujo social de la información, que a su materia estrictamente.

Se han analizado los problemas de ruido derivados de la ambigüedad con que pueden ser interpretados los descriptores dependiendo del contexto. Se han seleccionado expresamente un tipo de descriptores en los que esta ambigüedad era conocida. Se trata de descriptores que en el contexto de las bibliotecas son susceptibles de ser interpretados dualmente según lo que son o según lo que tratan. Almanagues ha sido el ejemplo paradigmático.

Los resultados para búsquedas en las que la correspondencia de la materia del documento con la búsqueda es la condición de su pertinencia, y con respecto al subconjunto de documentos indizados con los descriptores seleccionados, han sido en el modelo A de una tasa de pertinencia media de 0,27 por ciento frente a un 100% del descriptor almanagues seleccionado como representativo de este tipo de descriptores en el modelo B.

Los resultados que con los mismos descriptores el criterio de pertinencia era el contrario es decir su tipo, han sido para el modelo A del 73 por ciento frente de nuevo el 100 por cien para los obtenidos con el descriptor almanagues aplicado según el modelo B.

Estos resultados, si bien han de ser corroborados en estudios más amplios parecen indicar claramente el camino a seguir para avanzar hacia sistemas de recuperación de información y documentos más eficaces que los actuales en los que los problemas de sobrecarga informativa, de "infoxicación" no han sido resueltos satisfactoriamente a pesar de los enormes avances realizados en todos los campos en los últimos años.

Los sistemas de recuperación basados en clasificaciones y lenguajes documentales centrados en las materias de los documentos no resuelven adecuadamente, como se ha demostrado en este trabajo, las necesidades informativas de los usuarios, fundamentalmente a causa del exceso de ruido producido en el resultado de las búsquedas.

Una de las causas principales del ruido es la polisemia de los términos utilizados.

La capacidad contextualizadora de sistemas enfocados al tipo documental puede contribuir poderosamente, como creemos haber demostrado en este trabajo, a la desambiguación de las búsquedas y a la recuperación de los documentos de un modo más eficiente que los sistemas basados exclusivamente en la descripción y recuperación a partir de su materia.

Los encabezamientos género/forma pueden jugar un papel de primer orden en este planteamiento. Todo hace indicar que su implantación puede influir en la reducción del *ruido*¹⁷ documental de un modo notable.

¹⁷ Aunque el silencio no ha sido estudiado en este trabajo y que serán necesarias investigaciones que lo confirmen, nos inclinamos a pensar que en este aspecto los resultados pueden ser tan significativos o más que con el ruido.

Finalmente, el desarrollo completo de este nuevo modelo con la implementación de reglas de inferencia asociadas a gestores de bases de conocimientos y otros recursos de inteligencia artificial, que ya se están aplicando en campos como la clasificación de géneros *web*, abre nuevas y esperanzadoras perspectivas de poder ver en pocos años un gran salto en la resolución de uno de los problemas que tal y como expresaba Ortega y Gasset en 1935 ha angustiado al hombre de nuestro tiempo.

6. Referencias

- BADKE, W. (2010). Information overload? maybe not. *Online*, 34(5), 52-54.
- BAWDEN, David, y Lyn Robinson. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science* 35, April 2009, 180 -191.
- BIBLIOTECA DE CATALUNYA (2011). Codis Marc 655 (termes d'indexació -- gènere/forma):CM-070 [en línea]. En su: *Manual LEMAC*. Disponible en http://www.bnc.cat/sprof/normalitzacio/lemac/cm070_2008.doc. [Consultado: 22 en. 2011]
- BIBLIOTECA NACIONAL DE ESPAÑA (2011). *Bibliografía española* [en línea]. Disponible en: <<http://www.bne.es/es/Servicios/BibliografiaEspañola/>>. [Consultado 13 enero 2011].
- *Manual de autoridades* (2010) [en línea]. Disponible en: <http://www.bne.es/opencms/es/Servicios/NormasEstadares/Docs/Manual_de_Autoridades.pdf>. [Consultado 22 enero 2011].
- *Fichero de autoridades* (s.d.)[en línea]. Disponible en: <http://catalogo.bne.es/uhtbin/authoritybrowse.cgi> [Consultado: 8 marzo 2011]
- *Términos Género Forma* (2010)[en línea]. En su: Manual de autoridades. Disponible en: http://www.bne.es/es/Servicios/NormasEstadares/ManualDeAutoridades/RegistrosDeAutoridades/Terminos_genero/. [Consultado 22 enero 2011].
- BIBLIOTECA VALENCIANA (2010). Inicio. [En línea]. Disponible en: <http://bv.gva.es/screens/historiabv_val.html>. [Consultado 10 diciembre 2010]
- CASTILLO BLASCO, Lourdes (2006) *Elaboración de un tesoro de información de actualidad y conversión en red semántica para su empleo en un sistema de recuperación periodístico* [en línea] Disponible en: http://www.tdx.cat/TDX/TDX_UV/TESIS/AVAILABLE/TDX-0103108-095658//castillo.pdf [consultado: 7 marzo 2011].
- CHOMSKY, N (1995). *The Minimalist Program*. MIT. Press. Cambridge Massachusetts, USA
- CLOEREN, H (1988). Søren Kierkegaard (1813-1855). En su: *Language and Thought*. Berlin: Walter de Gruyter, 1988.
- CORNELLA A. (2000) *¿Cómo sobrevivir a la infoxicación?* [En línea]. Disponible en: <http://www.infonomia.com/equipo/articulos/infoxicacion.pdf>

[Consultado: 7 dic. 2010]

CROWSTON, K. and B.H. Kwasnik. (2003) Can document-genre metadata improve information access to large digital collections? *Library Trends*, Fall 2003.

DÍAZ ARIAS, R. et al. (1986) "La base de información de actualidad (BASINFA) de RTVE, un sistema automatizado de documentación periodística". En: *Segundas Jornadas Españolas de Documentación Automatizada*. Torremolinos, pp. 175-183

ESPAÑA (1978). Constitución Española. Artículo 20. *Boletín Oficial del Estado*, Madrid 1978

ESPAÑA (1985) Ley 16/1985, de 25 de junio de Patrimonio Histórico Español. *Boletín Oficial del Estado* 1985, 155, p. 20342-20352.

GENETTE, Gerard (1979). *Introduction a l'architexte*. Paris: Seuil. p: 87

GIL LEIVA, I. (2008). *Manual de indización. Teoría y práctica*. Gijón: Trea, 2008

GUINCHAT, Claire y Menou, Michel. (1992) *Introducción general a las ciencias y técnicas de la Información y Documentación*. Madrid: CINDOC, 1992, p. 19

HIMMA, K. E. (2007). The concept of information overload: A preliminary step in understanding the nature of a harmful information-related condition. *Ethics and Information Technology*, Seattle, Springer 2007

LA FONTAINE, H. et Otlet P. (1895) *Sur la création d'un répertoire bibliographique universel*. Bruxelles: [s.n.], 1895

LAMARCA LAPUENTE, María Jesús. (2009). *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. [En línea].

Disponible en: <http://www.hipertexto.info/documentos/indizacion.htm>.

[Consultado: 7 dic. 2010]

LEVY. D. (2005) "To Grow in Wisdom: Vannevar Bush, Information Overload, and the Life of Leisure". *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 281-286. Denver, CO, 2005

LIBRARY OF CONGRESS (2011) *Genre/Form Headings at the Library of Congress* [en línea]. Disponible en:

<<http://www.loc.gov/catdir/cpsd/genreformgeneral.html>>. [Consultado 7 marzo 2011].

- LITERARY THEORY (2009) [en línea]. *Wikipedia*. Disponible en: http://en.wikipedia.org/wiki/Literary_theory. [Consultado: 8 marzo 2011]
- Lloyd Bitzer (1968). The Rhetorical Situation. *Philosophy and Rhetoric*, 1 (January, 1968), 1-14
- LÓPEZ YEPES, José (2002) Documentación En su: López Yepes, José (coord.) *Manual de ciencias de la documentación*. Madrid, Pirámide. P. 39-72
- MANIEZ, J. (1987). *Les langages documentaires et classificatoires: conception, construction et utilisation dans les systèmes documentaires*. Paris: Éditions de la Découverte.
- MCLUHAN, M. (1998) *La galaxia Gutenberg: Génesis del homo typographicus*. Barcelona: Círculo de Lectores
- MILLER, Carolyn. (1984) Genre as Social Action. *Quarterly Journal of Speech*. 70 (1984): 151-67.
- MONTESI, Michela. Géneros web: líneas de investigación. *El profesional de la Información*, 2008 septiembre-octubre, v. 17, n. 5, pp. 551-557.
- MORENO FERNÁNDEZ, Luís Miguel y María Dolores Borgoños Martínez (2002). *Teoría y práctica de la Clasificación Decimal Universal*. Madrid. Trea.
- MOSCOSO, Purificación. (2002) Sistemas de información documental: concepto, modelo, estructura y organización. En: López Yepes, José. *Manual de ciencias de la información*. Madrid: Pirámide, 2002, pp. 519-536.
- MWIHAKI, S. (2004) Meaning as Use: A Functional view of Semantics and Pragmatics. En *Swahili Forum*. 11: 127-129
- ORTEGA Y GASSET (1935). "Misión del bibliotecario". *Congreso Internacional de Bibliotecas y Bibliografía*. 2º, Madrid. 1935
- PICARDI, Eva y Carlo Penco (2001). *Frege, G. "sentido y referencia"*. Bari: Editori Laterza. 2001.
- PHILOSOPHY OF LANGUAGE (2009)[en línea]. En: *Wikipedia*. Disponible en http://en.wikipedia.org/wiki/Philosophy_of_language. [Consultado: 8 en. 2011]
- ROMANOS DE TIRATEL, Susana (2008) Paul Otlet, el antepasado olvidado. Revisión bibliográfica. *Información, Cultura y Sociedad*, 2008, 18, p.13-36
- SAUSSURE, Ferdinand de (1983 [1916]). *Course in General Linguistics*. Eds. Charles Bally and Albert Sechehaye. Trans. Roy Harris. La Salle, Illinois:

Open Court. 1983

SAHLY, Nancy. (1985) Data elements used in Archives, Manuscripts, and Records Repository Information Systems: A dictionary of Standard Terminology En: *Marc for Archives and Manuscripts: The AMC Format*. Chicago: Society of American Archivists.

STAINTON, Robert J. (1996). *Philosophical perspectives on language*. Peterborough, Ont., Broadview Press.

TOFFLER, A. (1970). *Future shock*. New York: Random House.

YI, Kwan (2006) *Desafíos de la clasificación automatizada utilizando sistemas de clasificación bibliotecarios* [en línea] Disponible en: http://archive.ifla.org/IV/ifla72/papers/097-Yi_trans-es.pdf [consultado: 8 marzo 2011]

7. Índices de figuras y tablas

Figura 1. Delimitación de tipos documentales en el formulario de búsqueda de la BNE	p. 26
Figura 2. Modelo A o clásico	p. 36
Figura 3. Entidades Modelo B	p. 37
Figura 4. Tipos documentales del modelo B	p. 13
Figura 5. Encabezamientos tipo 1	p. 41
Figura 6. Encabezamientos tipo 2	p. 42
Figura 7. Encabezamientos tipo 3	p. 43
Figura 8. Encabezamientos tipo 4	p. 44
Figura 9. Formulario de búsqueda en la BNE	p. 49
Figura 10. Documentos devueltos en la búsqueda	p. 49
Figura 11. Documento pertinente	p. 50
Figura 12. Búsqueda por tipo documental en hemeroteca BV	p. 51
Figura 13. Documentos devueltos en búsqueda de la fig. 12	p. 51
Figura 14. Búsqueda por materia en hemeroteca BV	p. 52
Figura 15. Documentos devueltos en la búsqueda de la fig. 14	p. 52
Figura 16. Resultados devueltos para almanaques en monografías	p. 54
Figura 17. Resultados para búsqueda 'anarquismo' en subformulario: 'prensa y revistas'	p. 56
Tabla 1. Tipos de encabezamientos	p. 45
Tabla 2. Comportamiento de los términos del modelo A	p. 48