



D. DAVID RODRÍGUEZ MATEOS, con D.N.I. N° : 52129809 X

AUTORIZA:

A que su tesis doctoral con el título: **“Aplicaciones de XML para la Documentación Periodística: Efectos Sobre los Centros de Documentación de Prensa”** pueda ser utilizada para fines de investigación por parte de la Universidad Carlos III de Madrid.

Getafe, 20 de abril de 2004

A handwritten signature in black ink, appearing to read "David Rodríguez".

Fdo.: David Rodríguez Mateos.

James Caird

~~Associação Brasileira~~

~~Associação Brasileira~~

~~Associação Brasileira~~

~~Associação Brasileira~~

H/TU 39
(2º Sotano)

Aplicaciones de XML para la documentación periodística: efectos sobre los centros de documentación de prensa

David Rodríguez Mateos

Universidad Carlos III de Madrid
Departamento de Biblioteconomía y Documentación

Director: Dr. Antonio Hernández Pérez

Tesis para la obtención del título de
Doctor en Documentación

Diciembre de 2003

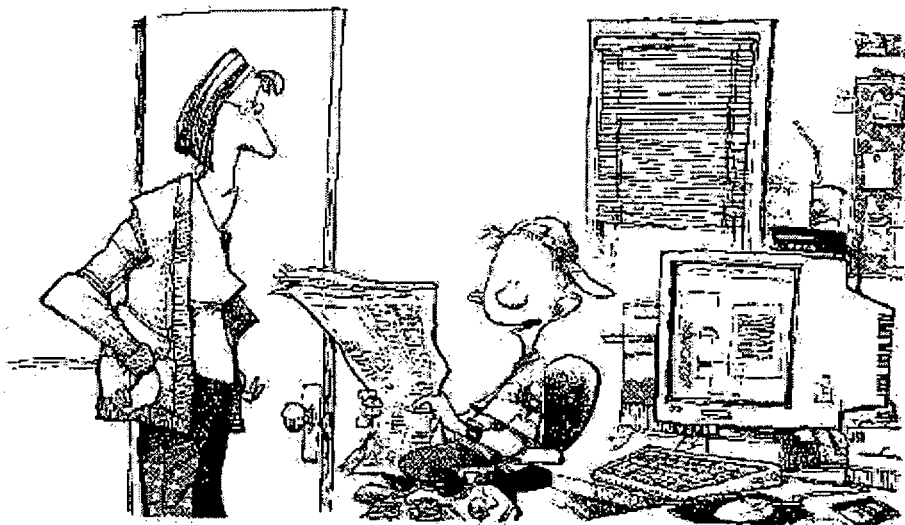


*Para mis padres, Juan y Juli,
y para mi hermano, Javier,
porque sigo creciendo con ellos,
cada día.*

*Para Félix,
por enseñarme el camino.*

*Para Tony,
por todo.*

*Para Ana,
por tener la fortuna de conocerla.*



Interesante... Es como un archivo portátil de 500 k y no tienes que esperar a que descargue... ¿Y dices que se llama "periódico"???

«Una cosa lamento: no saber lo que va a pasar. Abandonar el mundo en pleno movimiento, como en medio de un folletín. Yo creo que esta curiosidad por lo que suceda después de la muerte no existía antaño, o existía menos, en un mundo que no cambiaba apenas. Una confesión: pese a mi odio a la información, me gustaría poder levantarme de entre los muertos cada diez años, llegarme hasta un quiosco y comprar varios periódicos. No pediría nada más. Con mis periódicos bajo el brazo, pálido, rozando las paredes, regresaría al cementerio y leería los desastres del mundo antes de volverme a dormir, satisfecho, en el refugio tranquilizador de la tumba.»

LUIS BUÑUEL

Agradecimientos

Es un lugar común decir que toda tesis es un trabajo individual. Ésta no puede ser menos, ni en esa consideración ni en la certeza de que, siendo algo que solo uno puede hacer, no podría conseguirse sin la ayuda, al menos en mi caso, de muchísimas personas a las que es justo recordar cuando se alcanza al fin la meta.

En primer lugar, por la cercanía diaria, a prácticamente todos los compañeros del Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid, que han compartido conmigo muchas alegrías y algún que otro sinsabor, que han estado continuamente interesados en el transcurso de la tesis y que, sobre todo, me han animado para que llegara a buen puerto. Unos cuantos de ellos, además, me han apoyado de una manera especial, y así debe ser reconocido.

En primer lugar, debo agradecer la confianza que me han prestado personas como la vicerrectora de la Universidad y catedrática del Departamento, Mercedes Caridad, el actual Decano de la Facultad de Humanidades, Comunicación y Documentación, José Antonio Moreira, del Director del Departamento, Elías Sanz, y del director del Doctorado en Documentación, Miguel Ángel Marzal, que han estado continuamente pendientes de esta tesis y, más aún, de mi actividad profesional y personal en el Departamento, distinguiéndome con su atención.

También es necesario reconocer el esfuerzo de muchos de mis compañeros en la docencia durante estos años en la Universidad Carlos III, que, especialmente en los últimos dos años, me han ayudado hasta el punto de relevarme en parte, en mis tareas cotidianas para dedicarme con más intensidad a la tesis, y en más de una ocasión, en ofrecerme su ayuda generosa. Espero no olvidarme de nadie si cito en ese caso a Ricardo Eito, Belén Pérez, Pilar Azcárate, Mari Paz Martín-Pozuelo o, en años anteriores, Araceli García y Arturo Martín Vega.

Aún mayor interés han tenido quienes, por formar parte del grupo habitual de trabajo o por diversas circunstancias personales y profesionales, de una u otra manera, se han

convertido además en amigos que no solo se alegran con lo bueno, sino que también me sufren y me toleran. Debo empezar por Carmen Arellano (Menchu), con quien comparto despacho, actividades, un montón de experiencias inolvidables –varias de ellas, en el transcurso de esta tesis– una buena dosis de humor y una amistad que espero poder mantener por muchos años: esta tesis es también un poco suya. Lo mismo puedo decir de Tomás Nogales, que me ha ofrecido durante todo este tiempo su ingenio y su talante generoso conmigo; de Boni Martín, que sigue enseñándome tantas cosas en lo profesional pero, sobre todo, en lo personal; de Pilar Beltrán, que ofrece lo que tiene y lo que no tiene, y siempre lo hace con eficiencia; de Eva Méndez, que me distingue con su amistad a pesar de su agenda a veces desbordada; y de Carmen Jorge, con quien he disfrutado de buenos momentos y he compartido otros no tan gratos, que suelen ser los que más unen.

También es justo agradecer a otras personas con las que he compartido el día a día, que han acabado siendo amigos y que, a pesar de no tener ahora cerca, siguen siendo parte importante. No puedo dejar de mencionar a Elena Castro, a Paco Tosete y a Almiro Pérez: gracias por enseñarme parte de lo poco que sé y por regalarme vuestra sonrisa.

Asimismo, he de recordar a los compañeros del Dpto. de Ingeniería Telemática, y sobre todo a Carlos Delgado, Luis Sánchez, Vicente Luque y M^a Carmen Fernández, por compartir conmigo parte de su conocimiento, de su tiempo y, sobre todo, de su paciencia y amabilidad.

Con otras personas no he compartido actividad laboral pero si un mismo recinto de trabajo, y es justo agradecerles la atención particular y el cariño que me han dedicado, mucho más allá de lo profesional. En ese apartado, además de todos los ya citados, se encuentran Isabel Iribarren, Ana Morales –que fue mi primera compañera de despacho–, Jesús Robledano, Simone Frieiro, Sonia Sánchez, Jorge Morato, Coro Pichardo, Liliana Maspons, David García Hernán o Alberto Bustos. También le debo un recuerdo especial a M^a Antonia García Moreno, que creyó en mí cuando no era el momento propicio, y a Félix del Valle, que me guió por el sendero adecuado en una forma tal que espero poder recompensarle algún día.

No obstante, aunque a veces nos olvidemos de ello, la Universidad no es todo en nuestra vida, y también ha habido muchísima gente –lo cual me sigue sorprendiendo aún– interesada por mi persona y, aún más sorprendente, por mi trabajo. Quizá porque la tesis sea algo que acaba convirtiéndose en un parto –o eso cuentan– debo agradecer a quienes me honran con su amistad, en la cercanía y en la distancia. Dado que son, afortunadamente, muchos, espero que me podrán perdonar si los cito de carrerilla y si, con las prisas me olvido de alguno.

En la cercanía, cada vez más lejana por mis propios excesos de trabajo, están Vicente, José Luis, Mila, Mayca, Juan Ga, Óscar Reillo, Merce, Óscar Luceño, Beni, Javi Delfa, Mari Carmen, Javi Castillo, Susana, Ana y Javier. También he de recuperar el tiempo que no he dedicado a saber de Rosa, de Carlos, de Tony, de Ángel, de los compañeros de la carrera que son aún amigos a pesar del tiempo transcurrido (Bego, Goyo, M^a Ángeles, Luis, Lucía, Elena, Marcos, Raquel y Santi) y de aquellos que, desde hace aún más tiempo, mantienen el contacto: Elena, Teresa, Paloma y Beatriz, que han sufrido la desdicha de escuchar más de una queja sobre este trabajo.

En la distancia, otra mucha gente me ha ayudado a mantener el ánimo alto y a replantearme algunas cosas en la vida. Así ha ocurrido con Anna, Estrella, Cristina, Charo, Lola, Isa, Trini y Carmen. Gracias a todas por escucharme cuando tocaba.

Pero, sobre todo, he de agradecer especialmente todo su apoyo a una serie de personas fundamentales. Mis padres, Juan y Juli, y mi hermano, Javier (junto con mi cuñada, Maite), me han apoyado siempre, y es hora de que vean que sus esfuerzos no han sido en vano. En cuanto a mi director de tesis, Tony Hernández, no hay suficientes motivos para compensar su confianza en mí, contra viento y marea, y espero haber contribuido a que, al menos, no se vea defraudado por el resultado.

Finalmente, a Ana le debo haber tenido el empuje suficiente para subir una montaña gracias a su cariño y su fe: de la suya y de la mía. Espero seguir ganándome el privilegio de compartir parte de su vida.

Disculpe el lector si este monólogo le pareció largo: es la forma de mostrar la fortuna que tengo al contar con tanta gente en la vida.

ÍNDICE GENERAL

ÍNDICE GENERAL 1

CAPÍTULO 1 INTRODUCCIÓN 13

1.1.	UNA SOLUCIÓN PARA MUCHOS PROBLEMAS.....	15
1.1.1.	<i>La documentación periodística: desarrollo histórico y tecnológico.</i>	19
1.1.2.	<i>Los cambios producidos por el periodismo digital en línea.....</i>	22
1.1.3.	<i>El factor clave: la posibilidad de reutilización</i>	24
1.1.4.	<i>Una nueva herramienta de representación documental: el marcado de texto.....</i>	25
1.1.4.1.	XML como herramienta documental periodística	27
1.2.	OBJETO DE ESTUDIO: CAMBIOS EN EL DOCUMENTO Y EN LOS CENTROS DE DOCUMENTACIÓN PERIODÍSTICA	28
1.3.	HIPÓTESIS: NUEVAS HERRAMIENTAS, NUEVAS POSIBILIDADES	29
1.4.	OBJETIVOS: MEJORAR LA DOCUMENTACIÓN PERIODÍSTICA DIGITAL	30
1.5.	JUSTIFICACIÓN.....	32
1.6.	RELEVANCIA DE LA PROPUESTA.....	34
1.7.	MARCO TEÓRICO.....	36
1.8.	METODOLOGÍA Y TIPO DE INVESTIGACIÓN	39
1.9.	ESTRUCTURA DE LA TESIS	41
1.10.	FUENTES EMPLEADAS Y REFERENCIAS BIBLIOGRÁFICAS	43
1.11.	LIMITACIONES	46

CAPÍTULO 2 LA DOCUMENTACIÓN PERIODÍSTICA 49

2.1.	INTRODUCCIÓN.....	51
2.2.	SOBRE EL CONCEPTO DE DOCUMENTACIÓN PERIODÍSTICA	51
2.2.1.	<i>Término, objeto y actividades.....</i>	51
2.2.2.	<i>Definición de documentación periodística</i>	55

2.3.	ETAPAS HISTÓRICAS DE LA DOCUMENTACIÓN PERIODÍSTICA	56
2.3.1.	<i>Antecedentes de la documentación periodística</i>	56
2.3.1.1.	El nacimiento del periodismo informativo. Los primeros esbozos de documentación periodística.....	58
2.3.2.	<i>Los primeros centros de documentación periodística</i>	60
2.3.2.1.	La toma de conciencia: el desarrollo en los años 20.	61
2.3.3.	<i>El trabajo documental en prensa: 1920-1945</i>	62
2.3.3.1.	El periodismo interpretativo	62
2.3.3.2.	Aumento de los servicios de documentación periodística. Primeras concepciones teóricas	63
2.3.4.	<i>El desarrollo documental a partir de 1945</i>	66
2.3.4.1.	Problemas técnicos y primeras soluciones	67
2.3.5.	<i>Aplicación de la informática a la documentación periodística</i>	69
2.3.5.1.	Las bases de datos	71
2.3.5.1.1.	Inconvenientes de las bases de datos	72
2.3.5.2.	Memorias ópticas y GED	73
2.3.6.	<i>El desarrollo de los centros de documentación en España</i>	74
2.3.7.	<i>El comienzo de la difusión periodística a través de Internet</i>	76
2.4.	ELEMENTOS DOCUMENTALES DE LA INFORMACIÓN PERIODÍSTICA.....	78
2.4.1.	<i>Características de los documentos periodísticos</i>	79
2.4.2.	<i>Tipos de usuarios de los documentos periodísticos</i>	83
2.4.3.	<i>Necesidades de información de los periodistas como usuarios de documentos periodísticos</i>	85
2.4.4.	<i>La documentación periodística como proceso: principios</i>	89
2.4.5.	<i>Funciones de la documentación periodística</i>	92
2.4.6.	<i>Actividades documentales sobre la información periodística</i>	96
2.4.6.1.	La entrada de los documentos	97
2.4.6.2.	El análisis formal	98
2.4.6.3.	El análisis de contenido	99
2.4.6.4.	La organización y conservación de los fondos.....	105

2.4.6.5.	La difusión documental: hacia una nueva visión a partir de los contenidos digitales.....	107
----------	--	-----

CAPÍTULO 3 EL PERIODISMO DIGITAL EN LÍNEA Y LA DOCUMENTACIÓN PERIODÍSTICA..... 117

3.1.	INTRODUCCIÓN.....	119
3.2.	VARIACIONES EN LAS CARACTERÍSTICAS DE LOS DOCUMENTOS PERIODÍSTICOS.....	120
3.2.1.	<i>Relación entre el aumento de la urgencia y las posibilidades de explicación.....</i>	<i>120</i>
3.2.2.	<i>El riesgo de la saturación.....</i>	<i>123</i>
3.3.	NUEVAS RESPUESTAS A LA DISPERSIÓN DOCUMENTAL.....	125
3.3.1.	<i>Estructuración de contenidos: documentos, piezas, objetos y seudopiezas.....</i>	<i>127</i>
3.3.2.	<i>Interrelación de documentos y componentes.....</i>	<i>133</i>
3.3.3.	<i>Reutilización de los contenidos documentales.....</i>	<i>136</i>
3.3.4.	<i>Integración del centro documental con otros sistemas del medio... </i>	<i>138</i>
3.3.5.	<i>Empleo de contenidos multimedia.....</i>	<i>141</i>
3.4.	MODIFICACIÓN DE LAS RELACIONES DEL CENTRO DE DOCUMENTACIÓN...	142
3.4.1.	<i>Aumento de la interactividad.....</i>	<i>142</i>
3.4.2.	<i>Autoservicio de los documentos por los usuarios.....</i>	<i>143</i>
3.4.3.	<i>Conservación: nuevos factores derivados de los documentos digitales.....</i>	<i>146</i>
3.5.	IMPLICACIONES ECONÓMICAS DE LOS CAMBIOS EN EL CENTRO DE DOCUMENTACIÓN.....	148
3.6.	INTERROGANTES ABIERTOS. EL USO DEL MARCADO DE TEXTO.....	153

CAPÍTULO 4 LENGUAJES DE MARCADO PERIODÍSTICO: USO GENÉRICO PARA LA DEFINICIÓN DE SIGNIFICADOS Y DE ESTRUCTURAS DOCUMENTALES..... 159

4.1.	EL MARCADO DE TEXTO: SÍNTESIS HISTÓRICA	161
4.1.1.	<i>XML: de un metalenguaje a una familia de lenguajes de marcado</i> .	162
4.2.	DESCRIPCIÓN DE ESTRUCTURAS Y SIGNIFICADOS MEDIANTE MARCADO DE TEXTO.....	164
4.2.1.	<i>Algunas disyuntivas del análisis de contenido: autores y modos de descripción</i>	166
4.2.2.	<i>Relación entre la descripción semántica y la descripción estructural: modelos de marcado</i>	168
4.2.3.	<i>Implicaciones de los modelos de marcado</i>	171
4.3.	EVOLUCIÓN HISTÓRICA DEL MARCADO DE TEXTOS PERIODÍSTICOS	177
4.3.1.	<i>Lenguajes de marcado semántico basados en metadatos</i>	182
4.3.2.	<i>Lenguajes definidos por IPTC: el uso de marcado semántico-estructural</i>	183
4.3.3.	<i>El futuro de las iniciativas de marcado periodístico: entre la competencia y la convergencia</i>	185

CAPÍTULO 5 DESARROLLOS CONCRETOS DE LENGUAJES

PARA MARCADO PERIODÍSTICO: DE RSS A SPORTSML	191
5.1. LENGUAJES BASADOS EN METADATOS.....	193
5.1.1. RSS.....	194
5.1.1.1. Breve sinopsis histórica: un lenguaje, dos versiones.....	194
5.1.1.2. Estructura básica y elementos opcionales	195
5.1.1.3. Rasgos documentales de RSS.....	196
5.1.2. PRISM.....	198
5.1.2.1. El lenguaje básico PRISM.....	200
5.1.2.2. Lenguajes documentales en PRISM.....	202
5.1.2.3. Otros lenguajes de marcado PRISM.....	203
5.1.2.4. Rasgos documentales de PRISM.....	205
5.2. LENGUAJES DE MARCADO SEMÁNTICO-ESTRUCTURAL	209

5.3.	NEWSML: MARCADO DE CONTENEDORES PERIODÍSTICOS	212
5.3.1.	<i>Estructura básica de los documentos NewsML</i>	216
5.3.2.	<i>Mecanismos de descripción en NewsML: empleo de lenguajes documentales</i>	220
5.3.2.1.	Lenguajes documentales en NewsML: el uso de clasificaciones IPTC	223
5.4.	RASGOS DOCUMENTALES DE NEWSML	225
5.4.1.	<i>Definición de estructuras documentales</i>	225
5.4.1.1.	La disyuntiva entre inclusión y referencia	225
5.4.1.2.	El concepto de pieza y el concepto de objeto en NewsML	228
5.4.2.	<i>Descripción semántica de contenidos en NewsML</i>	229
5.4.2.1.	Definición de mecanismos para contener lenguajes documentales mediante el formato NewsML	230
5.4.2.2.	Aplicación de metadatos en NewsML	233
5.4.2.2.1.	Aplicación de metadatos que contengan atributos con valores basados en lenguajes documentales	234
5.4.2.2.2.	Asignación de metadatos mediante propiedades y valores: el elemento Property	235
5.4.2.2.3.	Indicación física de la ubicación del lenguaje documental... ..	238
5.4.2.2.4.	Referencia a lenguajes documentales mediante punteros	239
5.4.2.2.5.	Atributos para la identificación de autores de metadatos	241
5.4.2.2.6.	Inserción de metadatos adicionales a los ya existentes	242
5.4.2.3.	Lenguajes documentales concretos definidos por IPTC: utilidades e interrogantes	242
5.4.2.3.1.	Flexibilidad en la aplicación de metadatos: atractivo para los medios	245
5.5.	NITF: DESCRIPCIÓN ESTRUCTURAL DE NOTICIAS TEXTUALES	245
5.5.1.	<i>Tipos de elementos: obligatorios y opcionales</i>	247

5.5.2.	<i>Combinación de los elementos en NITF mediante módulos flexibles</i>	250
5.5.3.	<i>Capacidad para la descripción detallada de términos: el texto enriquecido</i>	253
5.6.	RASGOS DOCUMENTALES DE NITF	255
5.6.1.	<i>Opciones para la descripción documental</i>	256
5.6.1.1.	<i>Descripción inserta en el documento</i>	257
5.6.2.	<i>Consideraciones sobre la precisión documental de NITF</i>	261
5.6.3.	<i>Hiperflexibilidad estructural de NITF: algunos interrogantes</i>	265
5.7.	SPORTSML	266
5.7.1.	<i>Un tema, múltiples variaciones</i>	267
5.7.2.	<i>Estructura básica de un documento SportsML</i>	268
5.7.3.	<i>Uso de lenguajes documentales en SportsML</i>	270
5.8.	RASGOS DOCUMENTALES DE SPORTSML	271

CAPÍTULO 6	PAUTAS PARA EL USO DOCUMENTAL DE LOS LENGUAJES DE MERCADO PERIODÍSTICO	279
6.1.	INTRODUCCIÓN	281
6.2.	RASGOS DOCUMENTALES COMPARABLES ENTRE LOS LENGUAJES ANALIZADOS	282
6.2.1.	<i>Resumen de rasgos comunes y diferenciadoras de los lenguajes analizados</i>	286
6.3.	IDENTIFICACIÓN E INTERRELACIÓN DE LOS COMPONENTES	289
6.4.	DESCRIPCIÓN DOCUMENTAL: POSIBILIDADES PARA EL ANÁLISIS DE CONTENIDO	295
6.4.1.	<i>Interrogantes sobre el uso semántico de NewsML</i>	297
6.4.2.	<i>El mercado semántico de seudopiezas, partes y fragmentos: capacidades ofrecidas por PAM y NITF</i>	298

6.4.3.	<i>Descripción mediante lenguajes de marcado individuales o mediante combinación de lenguajes</i>	300
6.5.	INTEGRACIÓN DEL MARCADO DOCUMENTAL EN EL FUNCIONAMIENTO DEL MEDIO	304
6.5.1.	<i>Hacia un único marcado para múltiples usos</i>	305
6.5.2.	<i>Redactores y documentalistas: diferentes autores de un mismo marcado.....</i>	307
6.5.3.	<i>Conexión entre el sistema de producción y el centro de documentación</i>	311
6.5.4.	<i>Nuevas posibilidades de interrelación y transformación de los documentos.....</i>	312
6.5.4.1.	Formas avanzadas de interrelación de documentos: el empleo de seudopiezas.....	314
6.5.4.2.	Utilidades de las relaciones entre documentos: algunos ejemplos.....	315
6.5.4.3.	Capacidades documentales adaptadas al tamaño de un medio	318

CAPÍTULO 7 UN MODELO DE APLICACIÓN DE MARCADO

DOCUMENTAL PERIODÍSTICO BASADO EN NEWSML Y NITF..... 325

7.1.	INTRODUCCIÓN	327
7.2.	ACEPTACIÓN DE NEWSML Y DE NITF COMO HERRAMIENTAS DE MARCADO PERIODÍSTICO	328
7.3.	REQUISITOS PARA EL USO INTEGRADO DE NEWSML Y DE NITF	329
7.4.	RESTRICCIONES APLICADAS A LA IDENTIFICACIÓN Y DESCRIPCIÓN MEDIANTE EL USO CONJUNTO DE NEWSML Y DE NITF	330
7.4.1.	<i>Posibilidades de uso de NewsML</i>	331
7.4.2.	<i>Definición de un documento periodístico mediante el uso conjunto de NewsML y de NITF.....</i>	333
7.4.2.1.	Relación entre documentos NewsML y objetos NITF	334

7.4.3.	<i>Simplificación de NewsML y NITF</i>	336
7.5.	ESTRUCTURA BÁSICA DEL MODELO PROPUESTO	337
7.6.	ESTRUCTURA INTERNA DE LOS FICHEROS NEWSML QUE CONTIENEN LENGUAJES DOCUMENTALES.....	339
7.6.1.	<i>Actualización y modificación de datos</i>	341
7.7.	ESTRUCTURA INTERNA DE LOS DOCUMENTOS PERIODÍSTICOS DESCRITOS MEDIANTE NEWSML	342
7.7.1.	<i>Identificación de los documentos</i>	343
7.7.2.	<i>Descripción semántica: el elemento NewsComponent como contenedor de metadatos</i>	344
7.7.3.	<i>Descripción estructural: el elemento NewsComponent como contenedor de estructuras</i>	346
7.7.3.1.	Documento periodístico que contiene una noticia simple.....	352
7.7.3.2.	Documento periodístico formado por dos objetos.....	353
7.7.3.3.	Documento periodístico formado por más de dos objetos, susceptibles de ser convertidos en piezas mediante distintas opciones.....	356
7.7.3.4.	Capacidades añadidas de interrelación de documentos y componentes	361
7.8.	ESTRUCTURA DE LOS OBJETOS TEXTUALES BASADOS EN NITF	363
7.8.1.	<i>Metadatos incluidos en un objeto NITF empleado en el modelo propuesto</i>	364
7.8.2.	<i>Elementos estructurales y de texto enriquecido de NITF empleados en este modelo</i>	366
7.8.3.	<i>Referencia a varios hechos dentro del mismo objeto. Posibilidad de seudopiezas</i>	369
7.9.	ALMACENAMIENTO: BASES DE DATOS CON CONTENIDO NEWSML Y NITF	373

7.9.1.	<i>Uso de NewsML y de NITF como herramienta de almacenamiento</i>	374
7.9.2.	<i>Acceso a las bases de datos y generación de otras nuevas</i>	375
7.10.	UTILIDADES DEL MODELO PROPUESTO	379
7.10.1.	<i>Reducción de los condicionantes lingüísticos en la recuperación de información</i>	379
7.10.2.	<i>Acción documental conjunta de redactores y documentalistas</i>	380
7.10.3.	<i>Fomento de la actividad informativa propia del centro de documentación</i>	381
7.10.4.	<i>Ampliación de las posibilidades de búsqueda</i>	383
7.10.5.	<i>Capacidades de autoservicio documental</i>	385
7.10.6.	<i>Nuevas fuentes de ingresos</i>	388
7.10.7.	<i>Posibilidades de manejo técnico de los contenidos</i>	389

CAPÍTULO 8 CONCLUSIONES Y NUEVAS VÍAS DE

INVESTIGACIÓN.....	395	
8.1.	CONCLUSIONES SOBRE LA DOCUMENTACIÓN PERIODÍSTICA APLICADA A CONTENIDOS DIGITALES DIFUNDIDOS EN LÍNEA	397
8.2.	CONCLUSIONES SOBRE EL USO DE MARCADO DE TEXTO PERIODÍSTICO CON FINES DOCUMENTALES	399
8.3.	CONCLUSIONES SOBRE LA APLICACIÓN PRÁCTICA DEL MARCADO DE TEXTO PERIODÍSTICO CON FINES DOCUMENTALES	401
8.4.	NUEVAS VÍAS DE INVESTIGACIÓN	403
BIBLIOGRAFÍA	407	
DOCUMENTACIÓN.....	409	
DOCUMENTACIÓN PERIODÍSTICA	410	
ANÁLISIS DOCUMENTAL	424	
MARCADO DE TEXTO	428	
PERIODISMO DIGITAL	442	

MEDIOS DIGITALES Y BASES DE DATOS DE NOTICIAS	449
SISTEMAS DE GESTIÓN DE CONTENIDOS	450
RECUPERACIÓN DE LA INFORMACIÓN	455
OTRAS OBRAS CONSULTADAS	458
GLOSARIO.....	461
ÍNDICE DE FIGURAS	471
ANEXOS	481
ANEXO 1. ESTRUCTURA DE DOCUMENTO MARCADO MEDIANTE UN SUBCONJUNTO DE NEWSML, EMPLEADO EN EL MODELO PROPUESTO	483
ANEXO 2. ESTRUCTURA DE OBJETO MARCADO MEDIANTE UN SUBCONJUNTO DE NITF, EMPLEADO EN EL MODELO PROPUESTO	485
ANEXO 3. EJEMPLO DE DOCUMENTO NEWSML BASADO EN EL MODELO PROPUESTO	487
ANEXO 4. EJEMPLO DE OBJETOS NITF BASADOS EN EL MODELO PROPUESTO.....	493
ANEXO 5. LENGUAJE PARA INTERRELACIONAR DOCUMENTOS, BASADO EN LA ESPECIFICACIÓN PRISM, Y REALIZADO EN FORMATO NEWSML.....	499
ANEXO 6. EJEMPLO DE CATEGORÍA TEMÁTICA DEFINIDA POR EL IPTC EN LOS <i>SUBJECT CODES</i>	503
ANEXO 7. VALORES POSIBLES PARA EL ELEMENTO DE NEWSML ROLE.....	511
ÍNDICE DE MATERIAS.....	513

Capítulo 1

INTRODUCCIÓN

1.1. Una solución para muchos problemas

Los medios de comunicación digitales accesibles en línea a través de Internet ofrecen hoy mucha más información que sus equivalentes en papel o en los medios audiovisuales. Y además de ofrecer más información, y más actualizada, también ofrecen mejores explicaciones de los hechos que suceden; en muchos casos, mediante un notable trabajo de los centros de documentación de esos medios.

Pero esta explicación se refiere únicamente a ciertos fenómenos informativos: grandes acontecimientos, temas específicos, etc. Y además, en muchos casos, la explicación no se actualiza cuando han sucedido nuevos hechos que aportan una nueva visión, con lo que esa explicación puede quedar desfasada.

Un caso concreto: un informe sobre la plantilla de un equipo de fútbol realizado previamente al comienzo de un campeonato puede hacer referencia a jugadores dentro de ese equipo que, finalmente no jugaron con el mismo, por ejemplo, porque fueron traspasados a otro conjunto.

¿Cuál puede ser el motivo? El primer problema al que se enfrentan los centros de documentación de los medios periodísticos digitales es la **cantidad de información** que reciben. Estos centros han de procesar cada vez más documentos, pero además, deben **establecer las relaciones** entre los mismos: se les exige no solo que analicen un documento, sino que hagan referencias continuas a múltiples documentos relacionados.

Por ejemplo: si se habla de determinados personajes en una información, es probable que se pretenda además explicar quiénes son mediante sendas biografías de los mismos, y que se indique la relación documental mediante enlaces desde la información inicial a las mismas. Y también es posible que cada documento sufra una **actualización continua**, con la obligación de añadirle nuevos detalles.

Ello supone que, en muchos casos, un documento no sea ya una pieza informativa física, una noticia, sino que el **concepto de documento** incluya el conjunto de todos los textos relacionados con un mismo tema, incluidos físicamente en ficheros distintos que pueden ser producidos en momentos diferentes, y que estén relacionados entre sí, por lo cual pueden ser recuperables en su conjunto.

Más aún: puede suceder que un usuario solo quiera aquel componente o fracción de un documento original que hable del tema que le interese, mientras que el documento mencione varios temas incluso sin indicarlo de forma expresa. Otro ejemplo, muy habitual sobre todo en la prensa escrita, puede ser una noticia en la que se hable de tres acontecimientos cuya única relación consiste en que el periodista que los narra sea la misma persona, incluyendo en el titular principal solo el primer acontecimiento que refiere y sin sumario alguno que explique después el cambio de tema.

Además de las dificultades ya mencionadas para el centro de documentación, la identificación, descripción y actualización de esos documentos debe hacerse en el **menor tiempo posible**. El **acceso en línea** a los contenidos permite que, en cualquier momento del día se le reclame al centro de documentación respuestas en cuestión de minutos.

A todo ello debe añadirse que el **número y tipo de usuarios** que pueden necesitar información ha crecido enormemente. No se trata ya de que los periodistas reclamen a los documentalistas de un medio cualquier dato necesario para realizar su trabajo. Ahora, al poder acceder a través de Internet, cualquier usuario puede estar interesado, dentro o fuera del medio, por cualquier información publicada y almacenada por el mismo, hasta el punto de pagar por el acceso a estos contenidos. Y, por lo tanto, demandará una respuesta a su petición no solo de inmediato, sino de un modo directo, sin tener necesidad de consultar específicamente al centro de documentación.

Finalmente, un documento puede ser una suma de distintos objetos con distintos formatos: ya contengan texto o información sonora o audiovisual. Los documentos periodísticos son hoy, en muchos casos, **multimedia**.

¿Cómo puede responder el centro de documentación a todas estas necesidades? Una primera alternativa consiste en la elaboración manual de nuevos productos que relacionen las distintas informaciones entre sí, como los citados dossiers. Pero, como puede entenderse, la cantidad de posibles peticiones de los usuarios y de posibles relaciones que haya que realizar para responder a esas peticiones, de modo manual, es sencillamente imposible.

Dado que se dispone de documentos digitales, una segunda alternativa de los medios periodísticos digitales ha sido el uso de buscadores, al estilo de los empleados en Internet. Es decir, se han introducido todos los documentos producidos por un medio periodístico en una base de datos y se le ha aplicado un motor de búsqueda similar al empleado para recuperar documentos en Internet.

Ninguna de las dos soluciones ha sido válida hasta ahora. La elaboración manual de los dossiers corre el riesgo de producir silencio —no se analizan documentalmente todos los temas, y aquellos que son analizados exigen que alguien se encargue continuamente de ellos, con el consiguiente riesgo de desfase—.

Pero su posible solución, el uso de buscadores que recuperen documentos, permite usar en la búsqueda solo palabras que aparezcan en el texto, lo cual produce ruido: muchos de los documentos que se encuentran al buscar por una palabra, o incluso por una frase, en el buscador de un periódico digital no tienen relación con lo que el usuario está buscando.

¿Cuál puede ser la posible solución? Plantear el problema a priori, es decir, establecer una forma de producir los documentos periodísticos que cumpla con estos requisitos:

- Facilitar el análisis documental de los mismos, el detalle de los temas, personajes, lugares, etc.
- Facilitar la descripción en detalle de un documento, pero también de cada uno de sus componentes, con el fin de que cualquier usuario pueda

recuperar solo aquella información que sea de su interés, por reducida que sea, sin tener que obtener información añadida que no ha demandado.

- Permitir establecer del modo más fácil posible relaciones entre los contenidos: dentro de un mismo documento y entre documentos distintos
- Hacer posible todo ello con el menor esfuerzo posible por parte del centro de documentación; al menos, en lo que al análisis formal se refiere.
- Permitir usar toda esa información, sobre el contenido de un documento y sobre sus componentes, para poder recuperar la información de todos los modos posibles, como respuesta a las más variadas formas de búsqueda.

La presente tesis pretende ofrecer una propuesta para resolver esas necesidades empleando la representación documental de contenidos periodísticos mediante el marcado de texto, y en concreto, mediante lenguajes de marcado definidos expresamente para representar esos contenidos periodísticos, definidos por instituciones de las que forman algunos de los más importantes medios del mundo.

El uso del marcado de texto, y de herramientas que utilicen esta forma de representar los documentos, puede conseguir que los contenidos producidos por un medio informativo digital sean utilizables documentalmente con eficacia desde el mismo momento de elaboración de los mismos. Para conseguirlo, debe responderse a las siguientes preguntas:

- ¿Cómo puede evolucionar el funcionamiento del centro de documentación de un medio con el fin de responder, en su medida, a las necesidades tradicionales y a todos los nuevos requerimientos que se han planteado o que se puedan plantear en un futuro?
- ¿Cómo puede aprovechar las posibilidades que le brinda el uso del marcado del texto y, en concreto, los lenguajes que se analizarán en este trabajo, para responder a esa evolución?

- ¿Qué ventajas, qué problemas y qué requerimientos se plantearían con el uso de marcado de texto con fines documentales, tanto en el funcionamiento del centro de documentación como de los otros sistemas del medio?

La respuesta a esas preguntas supone en buena medida el replanteamiento de la actividad documental periodística: se requiere no solo un uso de nuevas herramientas técnicas, sino asumir la idea de que la actividad documental no sea posterior a la elaboración periodística, de que haya de tenerse en cuenta la utilidad documental de cualquier contenido producido en un medio desde el mismo momento de su elaboración.

Nuestra propuesta no es únicamente la elección de una forma técnica de representación documental de contenidos, el marcado de texto, sino el análisis sobre cómo puede emplearse esa herramienta basándose en un punto de vista documental que sea compatible al tiempo con el funcionamiento del resto del medio y, más aún, que potencie el funcionamiento de todo su conjunto.

¿Provocará, pues, el uso del marcado de texto para representar contenidos periodísticos accesibles en línea una nueva transformación de la documentación periodística y de su papel dentro de los medios de comunicación?

En realidad, el influjo de la tecnología sobre la actividad de los centros de documentación periodística no es nuevo. Observando la historia de la disciplina, podremos deducir que su evolución ha ido pareja al desarrollo de la actividad periodística y de la tecnología existente en cada momento.

1.1.1. La documentación periodística: desarrollo histórico y tecnológico

En origen, los centros de documentación de prensa eran meramente depósitos denominados a menudo *morgues* (con el cariz que, inconscientemente, se le daba a

esos centros como “cosa muerta”) y, posteriormente, como *archivos de prensa*, que contenían los fondos del propio medio, con un valor fundamentalmente histórico.

A medida que crecieron los fondos de los periódicos, y que dejaron de ser meros depósitos para comenzar a ser consultados con mayor frecuencia, comenzó a surgir la necesidad de contar con herramientas de descripción que permitieran localizarlos. Así surgieron los *índices de artículos*, primero más sencillos, como los alfabéticos o los temporales, y posteriormente, con cierto grado de análisis documental, como los onomásticos o los temáticos.

Poco a poco, se comenzó a valorar la importancia de introducir, mediante el uso de la documentación, explicaciones más allá de la mera mención de los hechos, para poder entender las informaciones en un mundo cada vez más complejo. Las dos grandes guerras mundiales del siglo XX produjeron las primeras muestras masivas de la información documentada.

Y entre medias, el periodo posterior de entreguerras dio origen al periodismo interpretativo, que pretendía analizar hechos sucedidos mediante el empleo masivo de información contextual recopilada por los documentalistas del medio. La Segunda Guerra Mundial provocó la asunción progresiva de ese periodismo, al menos, en Estados Unidos.

La segunda mitad del siglo provocó la extensión progresiva de esta práctica en diversos países, si bien muy poco a poco, dado que requería el empleo de recursos no siempre al alcance de todos los medios. La documentación periodística comenzó a ser considerada como una disciplina necesaria para realizar un periodismo de calidad.

Pero, a cambio, los centros se enfrentaron cada vez más a un exceso de información disponible, frente al cual comenzaron a hacer frente con dos armas: el intento de mejorar los documentos de descripción –los índices– y el empleo de herramientas tecnológicas, como la microfilmación de contenidos, cuyos costes a veces no compensaban la inversión.

La segunda herramienta tecnológica empleada fue la informática, desde comienzos de los 70, que facilitó las tareas cuantitativas de análisis y almacenamiento. Inicialmente, se tradujo en el uso de bases de datos referenciales, que no eran sino los mismos índices, organizados por campos. El empleo de ordenadores que facilitaban la búsqueda de términos concretos y fragmentos físicos de los mismos aumentó las posibilidades de búsqueda y redujo, en principio, el tiempo de recuperación de los datos.

Con el avance en los soportes de almacenamiento informáticos, surgieron las bases de datos a texto completo. Cada pieza periodística equivalía a un documento en la base de datos. Pero, en la mayoría de los casos, esa pieza estaba desgajada del resto de las informaciones junto a las que se publicó.

Los contenidos documentales de los periódicos dejaban así de tener meramente un valor histórico, sociológico, etc., para comenzar a tener un valor económico, dado que surgieron clientes interesados en el consumo masivo de esas informaciones. De todo ello se derivó una industria de difusión electrónica de la información: los medios comenzaron a obtener ingresos adicionales por sus fondos documentales y por el tratamiento de los mismos, en colaboración con empresas informáticas y telemáticas.

Con el tiempo, algunos medios comenzaron a comercializar por si mismos sus propios contenidos. Fue a comienzos de los 90, empleando los soportes ópticos, como el CD-ROM, cuando la difusión de los contenidos documentales por parte de los propios medios, sin ser masiva, sí comenzó a ser dirigida hacia el gran público. Esta solución facilitaba el almacenamiento, pero no la recuperación: había que esperar a que pasara un tiempo para obtener todos los fondos producidos por el medio en ese periodo.

1.1.2. Los cambios producidos por el periodismo digital en línea

¿Cuáles son, entonces, las innovaciones tecnológicas que inducen al uso del mercado de texto y que provocan un cambio en la concepción de la documentación periodística? Aquellas que previamente han producido un cambio en la propia concepción del periodismo.

La primera de ellas es el **periodismo en línea**, es decir, la posibilidad de ofrecer y de recibir contenidos periodísticos a través de redes telemáticas, es decir, de Internet.

La transmisión de contenidos periodísticos a través de redes telemáticas no surge con el uso de Internet, dado que ya se realizaba desde los años 70, entre las agencias de prensa y los medios periodísticos. Pero es Internet el medio que permite la difusión masiva de información en línea, que provoca el aumento de contenidos transmitidos y que, por lo tanto obliga aún más a organizar documentalmente esos contenidos para que el medio pueda ofrecérselos a múltiples usuarios distintos.

La segunda innovación tecnológica es el uso de **documentos digitales**. El hecho de emplear documentos digitales y de difundirlos en línea ha permitido que los medios hayan roto una serie de condicionantes asociados a la difusión de información periodística:

- Los **condicionantes espaciales**: cualquier persona puede acceder físicamente, de un modo sencillo y desde un ordenador, a cualquier contenido periodístico publicado en cualquier lugar.
- Los **condicionantes temporales de difusión**: el medio puede difundir cualquier contenido simultánea o inmediatamente al desarrollo de un acontecimiento.
- Los **condicionantes cuantitativos**: el medio puede difundir sus contenidos en cualquier cantidad: no tiene que contar con la necesidad de incluirlos en lotes, como las ediciones impresas con un número determinado de páginas.

- Los **condicionantes temporales de recepción**: los medios accesibles a través de Internet permiten el acceso a cualquier contenido en cualquier momento; pero además, en muchos casos, permiten además realizar instantáneamente una copia de esos contenidos, sin pérdida alguna de calidad.

A esta ruptura de las restricciones tradicionales se han añadido nuevas posibilidades de difusión:

- El empleo de los contenidos *multimedia*: en el mismo medio pueden incluirse textos con la profundidad empleada en la prensa impresa, imágenes en movimiento y sonidos.
- La capacidad de ofrecer múltiples mecanismos de *recepción* de información periodística. Además de la lectura secuencial, los receptores pueden acceder a los contenidos mediante la navegación a través de hiperenlaces y la recuperación de información en bases de datos. Y no son las únicas: la topografía semántica (los llamado *topic maps*) pueden ser un ejemplo de futuro, aunque no el único.
- La *interactividad* del nuevo medio. El medio informativo en línea permite que los usuarios puedan interactuar cuando reciban esas noticias: buscando más información, relacionándola con noticias similares, buscando diferentes modos de lectura, aportando su noticia o datos añadidos a la misma, etc.

Pero además, los propios periodistas de los medios también pueden interactuar con las noticias que ya han publicado: saber qué se ha publicado en su propio medio o en los medios a los que puedan tener acceso sin restricciones, poder elaborar e interrelacionar informaciones más complejas y, sobre todo, elaborarlas una sola vez y recuperarlas fácilmente, de modo que puedan liberar más tiempo para buscar y elaborar nueva información.

1.1.3. El factor clave: la posibilidad de reutilización

El centro de documentación de un medio periodístico posee, por tanto, una mayor cantidad de documentos, producidos digitalmente, accesibles y difundibles en línea físicamente divisibles y relacionables entre sí, que contienen diferentes tipos de información, pueden ser legibles de diferentes maneras. Con ese material, han de responder a unos usuarios que cada vez más reclaman acceder con mayor interactividad, pero de modos muy diferentes, a un mayor número de documentos para obtener una información de mejor calidad, en cualquier momento y sin la intervención directa del centro de documentación a la hora de recuperar los contenidos.

¿Cómo integrar los contenidos documentales existentes con las peticiones crecientes de los usuarios? El concepto clave para responder a esta pregunta es la **reutilización**, la capacidad para que cualquier contenido digital pueda ser utilizado cuantas veces sea necesario, para cualquier fin.

Pero no solo habrá de conseguirse que los contenidos sean reutilizables. Además, el centro de documentación deberá tender a que sean reutilizables otros dos tipos de información relacionada con los contenidos:

- Las **identificaciones**, o informaciones sobre dónde está físicamente cada uno de los documentos y cada componente de los mismos. Ello permitirá que, cuando se pretenda relacionar físicamente un documento, o un componente del mismo, con otro, pueda hacerse del modo más fácil posible, incluso de un modo automático.
- Las **descripciones**, es decir, la información que explica cual es el contenido de un documento: qué temas trata, a qué personajes hace referencia, en qué lugares transcurre la acción narrada, qué conceptos son tratados en el texto, etc.

Si se consigue que un único análisis documental incluya identificaciones y descripciones de contenidos reutilizables, se conseguirá que esos contenidos también lo sean. Un ejemplo: si se realiza una descripción precisa de los contenidos incluidos en un documento, el sistema podrá automáticamente relacionar, con mayor precisión, ese documentos con otros muchos.

Pero, al tiempo, si también se consigue detallar cada uno de los componentes de un documento, será más fácil localizarlos por separado en una recuperación posterior, y será más factible usarlos para responder a cualquier demanda: ya sea un periodista, que quiera complementar un hecho reclamando información similar producida anteriormente; un miembro del equipo de producción, que desee recuperar reportajes sobre un tema para no repetir el mismo planteamiento; un gestor del medio que quiera localizar cuántas informaciones ha realizado un colaborador del medio para pagarle sus servicios; o un usuario, que desee encontrar todas las entrevistas realizadas a su cantante favorito.

1.1.4. Una nueva herramienta de representación documental: el marcado de texto

Por lo tanto, una posible respuesta a todas estas necesidades sería el empleo de una herramienta de representación documental que permitiera realizar un único análisis documental muy preciso que identificara y describiera con precisión los fondos producidos por un medio de información, con el fin de facilitar su reutilización.

Pero debe contarse con que sea muy posible que no haya tiempo material para realizar un análisis como se ha realizado tradicionalmente hasta ahora. Es preciso, pues, que ese análisis se realice de otra forma y en otro momento. O al menos, que los documentos ya lleguen al centro de documentación habiendo recibido al menos un análisis documental básico.

¿Significa eso que los periodistas habrán de convertirse en documentalistas? No necesariamente. Pero sí significa que los periodistas participarán, probablemente sin

darse cuenta, en el tratamiento documental que permita la reutilización futura de contenidos desde su misma elaboración.

El modo más flexible de mantener el control documental sobre esas informaciones consiste en introducir en las mismas un *marcado de texto*, tanto de partes generales como de aspectos más concretos. En particular, ese marcado debe permitir tres tareas principales:

- Señalar aquellas partes que, estructural o semánticamente, tienen un significado tanto por sí mismas como en cuanto identificadores o puntos de acceso a esa información.
- Poder añadir etiquetas adicionales, denominadas metadatos, al margen del texto, que permitan añadir descriptores documentales de todo tipo (temáticos, funcionales, etc.), comunes al conjunto de la información.
- Permitir que la elaboración de ese marcado sea compartida, que en él puedan intervenir tanto los redactores que elaboran la información como los documentalistas que la analicen para que sea recuperable en el futuro.

Los contenidos marcados permitirán alimentar todo tipo de bases de datos que permitirán la recuperación de esas informaciones, con el fin de facilitar todo tipo de consultas, así como la elaboración de nuevos productos informativos.

Además, sin necesidad de modificar los contenidos almacenados, el marcado de texto hace posible también establecer todo tipo de relaciones, tanto hipertextuales como de otros tipos, que permitan la navegación a través de las informaciones periodísticas, ya sea en su redacción original o en los nuevos productos informativo-documentales que puedan producirse. Estos últimos, a su vez, no necesitarán de una reelaboración, sino que podrán componerse a partir de referencias a documentos ya elaborados o incluso a partes de los mismos.

1.1.4.1. XML como herramienta documental periodística

La respuesta a todos estos retos puede estar fundamentada en el uso de una herramienta tecnológica concreta: XML. Se trata de un metalenguaje de marcado de texto, dado que permite a su vez elaborar otros lenguajes específicos para diversos usos. En el caso del marcado de texto con fines periodísticos, algunos de esos lenguajes son los definidos fundamentalmente por dos consorcios.

El primero de ellos es el IPTC, que ha creado una familia de lenguajes para la definición estructural de conjuntos de contenidos periodísticos, como es el caso del lenguaje NewsML; para la definición de las partes de una noticia textual y para la descripción de la misma hasta sus más mínimos detalles, mediante el lenguaje NITF; y para la descripción hasta el más mínimo detalle de contenidos deportivos, mediante el lenguaje SportsML.

El segundo consorcio dedicado a la definición de este tipo de lenguajes es IdeAlliance, que está elaborando la iniciativa PRISM, consistente en un lenguaje básico de descripción de metadatos periodísticos genéricos, de igual nombre, que puede ser empleado junto a otros lenguajes para la elaboración de noticias (PAM), la inserción de descripciones de detalles concretos (PIM), la definición de relaciones entre documentos, etc.

Al margen de estos consorcios, se ha desarrollado una tercera alternativa, RSS, lenguaje de descripción básica de contenidos periodísticos mediante metadatos. RSS está actualmente en fase de desarrollo de una nueva iniciativa, basada en este lenguaje, pero dirigida hacia una ampliación de sus capacidades, limitadas en comparación a los lenguajes antes citados.

1.2. Objeto de estudio: cambios en el documento y en los centros de documentación periodística

Esta tesis pretende analizar cómo la producción de los nuevos documentos digitales difundibles en línea y sus respectivos usos por parte de un medio periodístico están influyendo en el funcionamiento de los centros de documentación de medios periodísticos. Y para ello, previamente, se analiza cuál ha sido la transformación en esos nuevos documentos producidos digitalmente y difundidos en línea a cualquier lugar del mundo, mediante Internet.

En concreto, se toman en consideración nuevos fenómenos como los siguientes: más contenidos a tratar, la necesidad de una mejor relación entre el centro de documentación de un medio y la redacción del mismo, el aumento de la urgencia en la demanda de resultados al centro, la utilización de herramientas informáticas que tengan en cuenta aspectos documentales, el requerimiento de acceso directo a los contenidos por los usuarios, así como el aumento de la consideración periodística y económica de los fondos documentales.

La alternativa ofrecida incluye la representación de esos contenidos mediante un marcado de texto empleado simultáneamente por todos los sistemas de un medio periodístico (no solo el centro de documentación, sino también la redacción, el sistema de producción y el sistema de gestión: administración, recursos humanos, contabilidad, marketing, planificación..., etc.) que facilite el análisis documental, reduzca al tiempo el esfuerzo material de todos esos sistemas, permita la representación más compleja de los contenidos, aumente las posibilidades de recuperación de los mismos y facilite el uso múltiple de éstos a la medida de cada posible usuario. El empleo de este marcado requiere el estudio de la influencia que puede tener su uso en el funcionamiento tanto de un centro de documentación como del resto de sistemas del medio.

Todas estas cuestiones provocan también la necesidad de analizar aspectos como:

- Los cambios en el concepto de documento periodístico digital y de las diversas partes que lo componen.
- La definición de las posibilidades genéricas del uso del marcado de texto para reflejar tanto las estructuras de los documentos periodísticos como para representar documentalmente sus contenidos.
- La aplicación de esas posibilidades en los lenguajes de marcado periodístico ya existentes.

De todo ello, se deducirá el desarrollo de un modelo concreto de aplicación que sirva como ejemplo para justificar la propuesta realizada.

1.3. Hipótesis: nuevas herramientas, nuevas posibilidades

La investigación realizada ha partido de las siguientes hipótesis:

1. El uso de documentos periodísticos digitales y su difusión a través de Internet ha aumentado a su vez el empleo de la documentación periodística, tomando como parámetros de medida el número de usuarios y la consideración por parte de los medios periodísticos sobre el valor de la misma.
2. El aumento de las ofertas de productos documentales por parte de los medios se ha producido en muchos casos mediante herramientas ajenas al ámbito documental, como los buscadores; y en los casos en los que sí ha habido tratamiento documental, éste no ha aprovechado todas las capacidades derivadas del hecho de manejar documentos digitales.
3. La existencia de lenguajes de marcado de texto periodístico, basados en XML y elaborados por instituciones que agrupan a entidades periodísticas, ofrece la posibilidad de asumir en un futuro esas herramientas en el seno de medios periodísticos.



4. Los lenguajes citados disponen de capacidades de tratamiento documental de los contenidos representados, tanto por la propia definición de los elementos que componen estos lenguajes como por el hecho de estar basados en XML y, asimismo, por la combinación de ambos factores.
5. El empleo de este tratamiento documental por parte del centro de documentación de un medio permitirá coordinar su actividad con la de otras secciones de éste, como la redacción, la producción o la administración del mismo.
6. La aplicación de un mayor tratamiento en profundidad de los contenidos documentales de un medio ofrecerá mayores posibilidades de representación y recuperación de los mismos, mejores capacidades de aprovechamiento de los contenidos por parte de todo tipo de usuarios y, por último, facilitará la elaboración de estrategias orientadas a la rentabilización de la información documental en un grado mayor al obtenido hasta ahora.
7. El empleo de soluciones basadas en XML y en lenguajes con visos de ser aceptados por multitud de medios en todo el mundo posibilitará el intercambio entre los mismos tanto de sus contenidos como de información documental sobre éstos.

1.4. Objetivos: mejorar la documentación periodística digital

Con esta investigación se pretenden cumplir los siguientes objetivos generales:

1. Responder a los requisitos que la difusión en línea de contenidos periodísticos digitales plantea a los centros de documentación de los medios, aprovechando esos requisitos no solo para ofrecer soluciones adecuadas, sino para mejorar el funcionamiento interno del propio centro y del medio que le acoge.

2. Mostrar como la documentación periodística, partiendo de su concepción tradicional, puede emplear las nuevas capacidades de los contenidos digitales para continuar mejorando la calidad de la información almacenada por el centro de documentación, pese a que éste reciba un mayor número de peticiones en calidad y en cantidad.
3. Establecer cómo la actividad de los centros de documentación periodística puede variar con el uso adecuado de una herramienta tecnológica, teniendo en cuenta no solo el mero planteamiento técnico, sino razonando su utilización desde el punto de vista documental, de modo que mejore la consideración del propio centro y, al mismo tiempo, el funcionamiento de todo el medio en su conjunto.
4. Proponer cambios específicos en la labor documental que reduzcan el empleo de recursos personales para tareas rutinarias, facilitando que los centros de documentación minimicen el tiempo dedicado al análisis formal de los contenidos, y mejorando la implicación de los documentalistas en el análisis de contenido y en la elaboración futura de contenidos periodísticos.

Todos ellos se desglosan concretamente en los siguientes objetivos específicos:

Determinar en profundidad cuáles son los nuevos cambios, producidos en los documentos digitales periodísticos difundidos en línea, que afectan a las características tradicionales de la documentación periodística.

Exponer cuáles han sido las modificaciones en las relaciones entre los centros de documentación de medios y sus usuarios, ya sean internos al medio o externos al mismo –particulares o corporativos–.

Establecer todas las posibles estructuras y componentes de los documentos que permitan identificar todas las partes de un documento, cualquiera que sea el tamaño de cada una de ellas, que puedan ser utilizadas en un futuro, así como los conjuntos de información formados por contenidos que, a pesar de haber podido ser publicados en diferentes documentos y fechas, tienen una relación periodística entre sí.

Ofrecer las capacidades del marcado de texto basado en XML y, en su caso, de las tecnologías concretas asumidas y asumibles por los propios medios, como herramienta práctica de representación documental que cumpla los requisitos anteriores.

Describir la evolución histórica de los lenguajes de marcado periodístico, y cómo el desarrollo de los mismos ha supuesto, de modo explícito o implícito, el aumento de la capacidad de descripción documental de los contenidos incluidos en los documentos producidos con esos lenguajes.

Analizar los principales lenguajes de marcado de texto con fines periodísticos actualmente existentes, delimitar sus capacidades documentales y mostrar cuáles de estas capacidades son más adecuadas para cumplir los objetivos anteriores, y cómo combinar las capacidades de esos lenguajes.

Establecer pautas que permitan comparar los citados lenguajes desde la perspectiva documental, así como proponer la combinación de los mismos, sin necesidad de modificar ninguna de las pautas básicas (ya sean las posibles estructuras de los elementos de cada lenguaje o el modo de combinar estos elementos), con el fin de respetar la posibilidad de difusión de contenidos documentales entre medios de cualquier tipo.

Proponer, basándose en el objetivo anterior, un modelo de aplicación de algunos de estos lenguajes, cuya elección se deberá a la capacidad de que los mismos sean aplicados por parte de un centro de documentación de un medio periodístico, en combinación con los otros sistemas del periódico.

1.5. Justificación

La presente tesis se justifica, además del interés objetivo de la misma, en la formación y el desarrollo académico del autor, así como en razones personales. En 1995, confluyeron una serie de hechos casi consecutivos: el final de la Licenciatura en Ciencias de la Información, en la rama de Periodismo; el surgimiento del interés

por la investigación en documentación periodística, fomentado por el Dr. Félix del Valle Gastaminza en los citados estudios, en la Universidad Complutense de Madrid; la introducción del autor en el uso de Internet, como usuario, y el surgimiento de la prensa accesible a través de Internet en España, a semejanza de lo sucedido ya en años anteriores en otros lugares.

La intersección de estas circunstancias produjo el interés por investigar la aplicación de la documentación periodística a esos nuevos contenidos periodísticos. El desarrollo de los mismos corrió paralelo a la formación posterior del autor, desde 1996, en los cursos del Doctorado en Documentación en la Universidad Carlos III de Madrid.

Muy pronto, contó con la tutela del profesor Antonio Hernández Pérez, del Departamento de Biblioteconomía y Documentación de la misma universidad, y licenciado y doctor en Ciencias de la Información por la Universidad Complutense de Madrid. A esta formación común a la del doctorado se unía también el interés por los mismos temas, así como en otros relacionados con las Tecnologías de la Información aplicadas documentalmente a contenidos accesibles en Internet.

Además de la dirección por el Dr. Hernández de la tesina de doctorado, esa tutela se manifestó mediante la inclusión en varios proyectos en los que participaron ambos, principalmente, dirigidos por la Dra. Mercedes Caridad Sebastián, del mismo Departamento. En esos proyectos, el autor participó como becario FPI del Ministerio de Educación y Ciencia, entre 1998 y 2000, bajo la tutela del Dr. Hernández, y desde octubre de 2000, como Ayudante de Escuela Universitaria en el citado Departamento.

Varios de esos proyectos estaban relacionados con el tratamiento de contenidos periodísticos representados mediante marcado de texto, dirigidos por el Dr. Carlos Delgado Kloos, del Dpto. de Ingeniería Telemática de la Universidad Carlos III de Madrid. Ese fue el caso del titulado *Periotrónico: Concepción y desarrollo de un periódico electrónico personalizado* (CICYT TEL97-0788), entre 1998 y 1999, y del subsiguiente proyecto *InfoMedia: un sistema multiplataforma para la publicación de*

información (CICYT TEL99-0207), entre 2000 y 2002, así como del proyecto *InfoFlex: Gestión de contenidos flexible y distribuida basada en tecnologías Web* (TIC2003-07208), dirigido por el Dr. Luis Sánchez Fernández, del mismo Departamento, que se desarrollará a partir de 2004.

En esos proyectos, se investiga el uso de lenguajes de marcado de texto periodístico basado en XML. De esa investigación se derivó el interés definitivo del director y del autor de esta tesis por desarrollar la aplicación de este metalenguaje, y de los lenguajes derivados del mismo desarrollados por instituciones internacionales, en los centros de documentación periodística pertenecientes a medios de comunicación accesibles en Internet.

Este interés por la documentación periodística y por su desarrollo en Internet se ha visto reflejado también en el plano docente, donde, tras participar como A.E.U. en diversas asignaturas relacionadas con las tecnologías de la información en la Diplomatura en Biblioteconomía y Documentación, el autor participa en la docencia de asignaturas, dentro de la Licenciatura en Periodismo de la Universidad Carlos III de Madrid, que están relacionadas con la documentación –como Documentación Informativa, desde el curso 2002/2003–, o con el uso de Internet con fines periodísticos, como Aplicaciones de la Red al Periodismo (desde el curso 2003/2004).

1.6. Relevancia de la propuesta

La relevancia de la solución propuesta puede medirse en términos periodísticos, económicos, de gestión y meramente documentales. Todos ellos se desglosan en los siguientes puntos:

- a) El desarrollo de una mejor oferta periodística que, con la llegada de Internet, se ha hecho mucho más compleja, en cuanto a sus posibilidades de producción, difusión y recepción, en respuesta además a un mundo cada vez más variable y complejo. Con esta oferta se pretende que los usuarios accedan no solo a una mayor cantidad de información, sino a unos

contenidos mucho más inteligibles, mediante el establecimiento de relaciones y otros mecanismos que permitan contextualizar mejor cada información publicada, tanto para uso de los propios generadores de la misma –redactores y, en menor medida, documentalistas– como para los futuros usuarios externos.

- b) El planteamiento por parte de los medios de una nueva forma de recepción de sus contenidos. Hasta ahora, los medios periodísticos se basaban en conceptos relacionados con la comunicación de masas (como la cantidad de audiencia). En adelante, habrán de plantearse la capacidad para atender a todo tipo de usuarios, a los que deben responder ofreciendo una mayor variedad en los medios, en los dispositivos físicos de salida (ordenador, papel, teléfono móvil, etc.) y en los modos de difusión: desde la mera petición de titulares hasta la oferta de informaciones con profundas contextualizaciones (análisis, informaciones complementarias, etc.), demandadas en un periodo cada vez más corto.
- c) La capacidad para mejorar la gestión de un medio. Esta mejora será visible por los periodistas, ya que sus redactores podrán utilizar con mayor eficiencia y eficacia los recursos ya producidos por el medio; pero también se percibirá desde un punto de vista económico, ya que se podrán plantear nuevas ofertas de contenidos atractivas para los posibles usuarios, en tantas formas como sea posible y en cualquier momento.

El esfuerzo que supone el replanteamiento en la elaboración de la información periodística aquí propuesto debe implicar la posibilidad de que tal iniciativa sea rentabilizada económicamente; o, al menos, en el caso de los medios que no planteen esa búsqueda de rentabilidad, que el modo de mejorar documentalmenete el funcionamiento documental del medio sea económicamente asumible.

- d) La mejora, desde el punto de vista documental, de los mecanismos de colaboración entre los distintos sistemas del medio periodístico –la

redacción, el centro de documentación, las áreas de producción y gestión económica y técnica– que permita optimizar el funcionamiento conjunto de todos ellos.

- e) La creación de las bases técnicas para que pueda mejorar la representación documental de los contenidos generados por un medio.
- f) La reducción del esfuerzo necesario para el tratamiento de la información periodística propia desde el mismo momento de su generación. Ello supondrá la búsqueda de formas de almacenamiento de esta información que permitan reutilizarla fácil y rápidamente.
- g) La potenciación del papel desempeñado por el centro de documentación dentro del propio medio, no solo desde un punto de vista meramente técnico, sino también con el fin de mejorar su imagen interna dentro del mismo.

1.7. Marco teórico

Dada la multiplicidad de temas tratados, el enfoque de esta tesis se ha realizado partiendo de disciplinas inicialmente dispares entre sí, pero que están interrelacionadas.

En primer lugar, la **documentación**, como disciplina de carácter instrumental. Más en concreto, este trabajo se enmarca en el ámbito de la documentación periodística, que pretende la mejor gestión de la información producida con fines periodísticos, mediante la definición de las estructuras de esta información y la aplicación de los correspondientes principios de descripción documental, almacenamiento, recuperación y difusión de la información.

El trabajo se centra en los contenidos documentales producidos por un medio periodístico en concreto, en su representación, en su descripción y en el planteamiento de múltiples opciones posibles de recuperación de esos contenidos.

Se ha hecho hincapié en la representación documental de contenido empleando una herramienta, el marcado de texto, que ofrece grandes capacidades para la descripción documental, tanto en su análisis formal como en el análisis de contenido de la estructura de los documentos y de los significados de los contenidos insertos en ellos.

Este texto representa un ejemplo sobre el modo en el que la mera información estructural puede, por sí misma, ofrecer información semántica, al distinguir cuales de los componentes de un documento pueden ser definidos como *unidades documentales* que puedan ser recuperadas por separado en un futuro.

Pero además, el trabajo asume no solo que los contenidos recibidos tienen una estructura determinada por sus productores –los redactores del medio–, sino planteando que la descripción documental puede llegar hasta el punto de definir nuevas estructuras, e incluso, de producir nuevos documentos no solo secundarios, sino también primarios.

En este sentido, la documentación periodística se plantearía no ya como una actividad que surge posteriormente a la elaboración de los contenidos, sino que comienza desde el mismo momento en el que esos documentos comienzan a elaborarse.

En segundo lugar, el **periodismo**, como disciplina sobre cuyo producto se desarrolla la actividad documental descrita en esta tesis. En particular, la tesis analiza, desde un punto de vista documental, el periodismo digital en línea, es decir, la producción digital de contenidos periodísticos que son accesibles a través de redes telemáticas, como Internet.

Este nuevo medio de producción y difusión supone notables cambios en el trabajo periodístico cotidiano. Algunos de esos cambios afectan al trabajo de los centros de documentación de los medios periodísticos, y desde esa perspectiva, son analizados. Esos cambios deben suponer una mejor relación entre la redacción de un medio y su centro de documentación.

En tercer lugar, más secundario, las **telecomunicaciones**, como técnica que propició el comienzo del mercado de texto aplicado a piezas periodísticas, con el fin de transmitir dichas noticias de unos medios a otros; y como técnica que continúa fomentando el desarrollo de este mercado.

Especialmente, se han tenido en cuenta los diversos desarrollos concretos, es decir, los lenguajes de marcado específicamente dedicados a representar medios periodísticos, que continúan fomentando instituciones como el IPTC, y que siguen lideradas en buena parte por expertos de las telecomunicaciones, más enfocados por definición hacia la transmisión física de contenidos.

El trabajo muestra no solo cómo unas herramientas definidas para la transmisión de contenidos tienen notables utilidades documentales sino también cómo los miembros de los consorcios dedicados a la definición de lenguajes de marcado periodísticos tienen cada vez más en cuenta planteamientos documentales.

Este aspecto se desarrolla cada vez con más fuerza mediante sistemas de clasificación de contenidos concretos, como el Sistema de Referencia Temática desarrollado y modificado por el IPTC, o como la adopción de herramientas de representación claramente documentales, como la iniciativa Dublin Core, adoptada como base en buena parte por lenguajes como los desarrollados en la iniciativa PRISM.

En cuarto y último lugar, la **informática**, como disciplina que permite el tratamiento automatizado de la información, interrelacionada con las disciplinas antes mencionadas. En particular, la relación con la documentación se establece mediante el empleo específico de los lenguajes de marcado.

En concreto, se utiliza como herramienta básica durante todo este trabajo el metalenguaje XML, subconjunto de SGML, una forma de establecer reglas y pautas genéricas comunes para la elaboración de lenguajes de marcado específicos, como los que se describen en este trabajo. XML ofrece la posibilidad de elaborar lenguajes que, partiendo de una definición de elementos comunes, puedan combinarse de una forma flexible para generar documentos distintos.

Esta forma de representación se adapta perfectamente a la generación de documentos periodísticos que, si bien presentan características comunes en los elementos y partes que los componen, ofrecen una gran variedad de combinación, por ejemplo, en la disposición y en el uso de estos elementos.

1.8. Metodología y tipo de investigación

La investigación descrita se ha realizado mediante una serie de fases en la que se ha combinado el método deductivo con el método inductivo. La primera de estas fases, de tipo deductivo, consiste en un análisis de la documentación periodística como proceso, con el fin de encontrar sus características comunes.

Una segunda fase, inicialmente deductiva, ha consistido en el análisis de la producción de documentos periodísticos digitales difundida en línea, de la cual se han extraído aquellos factores que han supuesto la modificación de algunas de las características documentales descritas anteriormente. Este análisis ha inducido a la definición de dos ideas clave: la necesidad de definir una división estructural del documento periodístico, teniendo en cuenta sus posibles componentes desde un punto de vista documental (el documento es divisible en piezas, objetos y seudopiezas), y el planteamiento sobre cuáles eran los interrogantes documentales a los que debía responder un centro de documentación de un medio periodístico para aprovechar correctamente la división anterior.

En una tercera fase, en este caso inductiva, se han tomado algunos conceptos básicos relativos al marcado de texto, y en concreto, a su formulación mediante XML, definiendo tres modelos para la descripción de contenidos mediante marcado de texto (inserto, adjunto y externo), teniendo en cuenta la posición del marcado respecto a la posición del contenido descrito, así como las implicaciones derivadas del uso de uno u otro mecanismo.

Una cuarta fase, en esta ocasión, deductiva, ha consistido en la descripción de cinco lenguajes concretos de marcado de texto periodístico, RSS, PRISM, NewsML, NITF y SportsML, que intercalan en su uso varios de los modelos descritos. Para cada uno

de ellos, se ha realizado una breve mención histórica contextualizadora de su nacimiento, se han descrito sus aspectos más destacables desde el punto de vista documental, y se han mostrado sus utilidades más concretas.

Tras esta descripción individual, se han comparado estos lenguajes entre sí, así como su uso combinado. Para ello, se han estudiado las capacidades de los lenguajes para identificar y para describir contenidos –haciendo hincapié tanto en el modo empleado para definir las estructuras de los documentos como en la forma de representar sus contenidos, ya fuera mediante lenguaje libre o mediante lenguaje controlado–, y al tiempo, para relacionar esos contenidos con otros, y para integrar el funcionamiento del centro de documentación con el de otras secciones del medio.

Una última fase, inductiva, ha consistido en la definición de un modelo de aplicación concreto para representar contenidos periodísticos mediante marcado de texto que respondiera a las necesidades planteadas y a las utilidades encontradas en los lenguajes de marcado. Se ha empleado como punto de partida el uso combinado de NewsML y NITF, reduciendo el número de elementos de cada lenguaje.

El requisito principal del modelo de aplicación consiste en no reducir la utilidad documental de ninguno de los lenguajes, sino, al contrario, mejorar la identificación y la descripción de los contenidos con el fin de hacer posible un mayor número de relaciones futuras entre los documentos marcados. Ello permitirá por lo tanto un mayor número de alternativas de recuperación de los contenidos, de lo que se derivan mayores posibilidades para la reutilización de los mismos.

En conjunto, el trabajo es en buena medida sincrónico, dado que el estudio se centra en la utilidad actual de los lenguajes de marcado estudiados, si bien su justificación ha requerido sendas introducciones históricas que contextualizaran la elección de las tecnologías estudiadas, tanto en el aspecto más teórico, el de su aplicación a la documentación periodística, como en el más pragmático, el desarrollo de las tecnologías analizadas.

Asimismo, se trata de un trabajo especulativo, que aporta una solución genérica a los problemas planteados mediante un modelo de aplicación susceptible de diferentes implementaciones concretas en el futuro.

1.9. Estructura de la tesis

El contenido de la presente tesis se desglosa en siete capítulos adicionales a éste. El capítulo 2 muestra de modo breve los fundamentos teóricos e históricos de la documentación periodística, con el fin de mostrar tanto las técnicas empleadas desde el surgimiento de la disciplina, en el siglo XIX, hasta nuestros días. Con este repaso, se ha pretendido enmarcar el contenido del trabajo tanto para aquellos especialistas en la materia como para aquellos lectores que, sin conocerla, puedan establecer un punto de partida bajo el cual les resulte comprensible el enfoque de los capítulos siguientes.

En el capítulo 3, se describe la relación entre el periodismo en línea, difundido sobre todo a través de Internet, y la documentación periodística, mostrando cómo ésta última se ha visto modificada mediante fenómenos como el aumento de la urgencia, las nuevas capacidades de respuesta a fenómenos como la dispersión informativa o las posibilidades de saturación, etc.

Estos cambios han provocado el surgimiento de nuevos factores (posibilidades añadidas de estructuración de los contenidos, capacidades de integración de la actividad documental en el funcionamiento del medio, etc.), así como la modificación de las relaciones del centro de documentación respecto a sus usuarios, incluyendo el aumento de la interactividad y de la capacidad del centro documental para ofrecer el autoservicio de sus fondos, así como una nueva toma en consideración del efecto de los usuarios sobre la conservación de los documentos.

De la muestra de los factores descritos en ese capítulo, se deduce la formación de una serie de interrogantes, a los cuales pretendemos responder mediante la aplicación del marcado de texto periodístico, desarrollado en dos capítulos posteriores. En el primero de ellos, el capítulo 4, se analiza el empleo del marcado de texto y, en

concreto, de XML: tras una breve sinopsis teórica e histórica, se plantean varios modelos genéricos de marcado (interno, adjunto y externo), y a continuación, se mencionan los cinco lenguajes concretos que gozan de aceptación, a diferentes niveles, en la comunidad científica y periodística internacional: RSS, PRISM, NewsML, NITF y SportsML.

Las posibilidades documentales de cada uno de estos lenguajes es descrita por separado –teniendo en cuenta que todos ellos asumen el uso de más de uno de los modelos genéricos antes empleados– en el capítulo 5. A continuación, en el capítulo 6, se ofrecen algunas pautas concretas de uso genérico de cada uno de estos lenguajes, se compara su funcionamiento, se muestran las ventajas en el uso de cada uno de ellos por separado y se describen las posibilidades de utilizar combinadamente dichos lenguajes de dos en dos. Asimismo, se reflexiona sobre las posibilidades que ofrece la combinación de ambos lenguajes para el trabajo integrado entre el centro de documentación, la redacción y el sistema de producción física del medio. También se incluye una mención genérica sobre las nuevas posibilidades que ofrece la combinación de NewsML y NITF.

En el capítulo 7, se propone un modelo de aplicación del marcado periodístico sobre contenidos textuales producidos por un mismo medio, empleando para ello la combinación de los dos lenguajes citados, cuya elección está basada en los argumentos desarrollados en capítulos anteriores. Se ofrecen una serie de pautas técnicas para su utilización y se aportarán posibles usos derivados de este modelo, mostrando además algunas posibles ventajas del mismo.

Finalmente, se mostrarán las oportunas conclusiones obtenidas de la presente investigación, y se propondrán nuevas líneas de trabajo derivadas de las mismas.

1.10. Fuentes empleadas y referencias bibliográficas

En cuanto a las fuentes empleadas para justificar las pautas seguidas en la investigación, han de destacarse en primer lugar, atendiendo al formato empleado, las fuentes electrónicas accesibles fundamentalmente a través de Internet. Ello se ha debido, especialmente en los aspectos más técnicos del trabajo, al hecho de que muchas de las materias tratadas presentan una continua evolución.

La difusión de este tipo de temas se realiza a través de documentos actualizados con frecuencia (en muchos casos, cuya renovación se realiza cada año, o incluso en lapsos menores) accesibles solo a través de determinados sitios web de las instituciones que los desarrollan. Ese ha sido el caso de los materiales empleados para la descripción de los lenguajes concretos analizados, cuya obtención ha partido de los sitios web del IPTC y del consorcio IdeaAlliance, por citar a sus principales promotores, aunque no los únicos.

No obstante, si bien los documentos concretos que rigen estos lenguajes varían –en algunos casos, aportando modificaciones de mayor o menor calado, pero habitualmente tendentes a mantener lo ya implementado– no varían tanto los sitios web correspondientes en los que se desarrollan estos lenguajes. En la bibliografía podrá encontrarse una división por soportes que permitirá la referencia concreta, en el caso de los sitios web, a aquellos que el lector pueda considerar de interés para complementar la visión sobre el tema, permanentemente en movimiento.

El segundo grupo de publicaciones en importancia lo constituyen las publicaciones periódicas, con especial atención a los artículos de revistas, muchos de cuyos textos también han sido obtenidos en formato electrónico a través de recursos como la colección de revistas de la Biblioteca de la Universidad Carlos III de Madrid, suscrita a texto completo a versiones digitales de publicaciones empleadas en la presente tesis, en particular, en los campos de la Informática y la Documentación; y en menor



medida, en versiones en papel de dichas revistas, especialmente en aquellos números de mayor antigüedad.

En éste último aspecto, debe destacarse la revista *Técnicas de prensa*, pionera no solo en la difusión del uso de XML aplicado al periodismo, sino también en la muestra de soluciones específicas realizadas por medios y grupos concretos. Ello no es de extrañar, dado que esta publicación es el órgano de difusión del consorcio internacional IFRA, dedicado a la aplicación de la informática y las telecomunicaciones al mundo de la prensa escrita.

Los contenidos de otras revistas, como *Cuadernos de documentación multimedia* o *Documentación de las Ciencias de la Información*, han sido útiles especialmente para la situación histórica de la documentación periodística en España.

El empleo de las monografías como fuentes de información ha estado enfocado mayoritariamente hacia aquellos epígrafes de la tesis, como las introducciones históricas y teóricas, que presentaban una menor oscilación en el tiempo, en cuanto a la duración en la validez de sus contenidos.

En este campo, deben destacarse entre otras las aportaciones de los profesores Gabriel Galdón, Antonio García Gutiérrez y Eulàlia Fuentes. Ésta última constituye una referencia imprescindible para el estudio de los centros de documentación periodística de los medios digitales en España en la última década, ya sea mediante sus obras en solitario o acompañada de los profesores Alfons González Quesada y Àngels Jiménez, que ha producido una tesis específica sobre el tema.

Una fuente de interés, especialmente en el último año, ha estado constituida por las publicaciones en congresos. En particular, dos han sido los ámbitos que hemos considerado de interés en los textos de este tipo que hemos consultado: la descripción sobre las actividades concretas de documentación periodística en distintos medios, por una parte. Y por otra, el desarrollo concreto de los lenguajes de marcado propuestos, que ha visto como aumentaba el número de comunicaciones destinadas a este fin en diversos lugares del mundo, si bien, aún tímidamente.

Asimismo, han sido de interés diversas publicaciones, accesibles en la red, referentes a aplicaciones de gestión de contenidos, elaboradas por sus propios fabricantes, que han servido para mostrar la cada vez mayor aceptación de la industria informática hacia los lenguajes propuestos –en particular, hacia NewsML y NITF–.

También han sido consultadas bases de datos referenciales y de texto completo de carácter informático, como Research Index, de carácter documental, como LISA, y de temática variada, entre las que destacamos a Scirus. Resulta reseñable el hecho de que, en ninguna de ellas, apenas se obtuviera información ni sobre el empleo documental del mercado de prensa periodístico ni sobre los propios lenguajes analizados.

El modelo seguido para las citas bibliográficas y las subsiguientes referencias ha sido el definido por la norma UNE 50-104-94, equivalente a la norma ISO 690:1987. En particular, para las citas bibliográficas, se ha optado por el método denominado de “primer elemento y fecha”, mientras para las referencias bibliográficas, basándose en la citada norma, se ha optado por realizar algunas variaciones definidas por Assumpció Estivill y Cristòbal Urbano a partir de la citada norma ISO.

Estas variaciones se deben a que la citada UNE 50-104-94 no contemplaba el empleo de fuentes electrónicas, las cuales, como se ha señalado, constituyen parte importante de las referencias empleadas en la presente tesis.

Para la elaboración de las citas y de las referencias bibliográficas se ha empleado el programa informático Procite 5.0, que permite su integración con el programa Word 2000, utilizado para la redacción del texto. La inserción de las referencias bibliográficas a las que remiten las citas se realiza al final del capítulo correspondiente, mientras que al final del contenido de la tesis se ofrecerá la bibliografía completa, incluyendo otras obras que, sin ser expresamente citadas, sí han formado parte de la investigación realizada y de otras relacionadas que puedan llevarse a cabo en el futuro.

1.11. Limitaciones

Aunque el objeto de la documentación periodística abarca todos los documentos periodísticos, ya sean textuales, visuales, sonoros o audiovisuales, este trabajo se referirá únicamente a los primeros. Ello se debe a que los documentos icónicos, sonoros y audiovisuales poseen ciertas características distintivas que exigen un estudio por separado, el cual excede el ámbito de este trabajo.

La tesis se enfocará, en particular, hacia un tipo de documentos periodísticos textuales; en concreto, sobre aquellos que se producen digitalmente en la redacción de un medio periodístico y que pueden ser difundibles en línea, a través de redes de ordenadores, sin perjuicio de que puedan realizarse copias de los mismos en otros soportes, como el papel.

Más concretamente, nos centraremos en el manejo por parte del centro de documentación de un medio informativo (o de un grupo mediático, en el caso de centralice la documentación de los contenidos por distintos medios de ese grupo) de aquellos documentos textuales periodísticos digitales que han sido o están siendo producidos por el mismo medio o grupo, haciendo hincapié en el uso de marcado de texto para representar estos documentos, y en las implicaciones de ese uso, especialmente en las múltiples capacidades de utilización de estos contenidos, una vez producidos periodísticamente.

Incluso delimitando la tesis a ese aspecto concreto, el abanico de posibilidades que ofrece la situación descrita en esta introducción, junto con el marco teórico en el que se encuadra obligan a recortar el ámbito en el que se encuadra el trabajo, descartando otros muchos temas como, entre otros, los planteados a continuación.

Uno de esas facetas es la repercusión documental de la oferta que muchos medios realizan sobre el acceso a fuentes primarias ajenas, sin recopilarlas en su propio medio. Por poner un ejemplo: el texto de un tratado internacional puede hoy incluirse

mediante una mera referencia hipertextual al documento incluida en el sitio web de la organización que lo promulga.

¿Cómo tener la certeza de que el contenido de ese documento pueda ser obtenido a medio plazo (dentro de algunos meses, por ejemplo)? Si la fuente primaria no garantiza el almacenamiento y recuperación de estos datos, ¿esa función debe ser asumida por el medio informativo? ¿Corresponde al medio garantizar el acceso a fuentes que aseguren ese acceso en el futuro? ¿Esas fuentes serán propias o ajenas, gratuitas o de pago? La respuesta a todas estas preguntas exige el desarrollo de un nuevo estudio al margen de la presente tesis.

Otro tema pendiente, sin agotar una lista que podría ser interminable, será la disyuntiva entre el análisis documental de contenidos realizado mediante su indexación automática de los contenidos por parte de programas informáticos que empleen de técnicas como la inteligencia artificial; o bien, mediante la indización por parte de especialistas en este tipo de análisis. Tanto las herramientas descritas como el modelo propuesto permiten el uso de cualquiera de las dos opciones.

Capítulo 2

LA DOCUMENTACIÓN PERIODÍSTICA

2.1. Introducción

El presente capítulo tiene por objeto definir, al menos de modo básico, el concepto de documentación periodística empleado en el presente trabajo. Esta definición es precisa dado que, comenzando con el propio concepto, ha recibido durante los últimos 25 años, diversas definiciones, divergentes en tres aspectos: el término usado, el ámbito concreto sobre el que se aplicaba y la actividad definida. Se hace necesario, por lo tanto, delimitar en los tres casos, siquiera brevemente, cuál es la opción elegida en el marco del presente trabajo.

A continuación, se realizará una descripción de las etapas históricas en las que la documentación periodística se ha desarrollado como disciplina hasta la actualidad. Esta descripción pretende mostrar como las necesidades actuales de esta disciplina derivan de otras anteriormente formuladas, y qué respuestas se les dio a aquellas, con el fin de partir de las mismas para resolver los interrogantes actuales.

En un epígrafe posterior, se mostrarán los elementos básicos de la documentación periodística como disciplina: sus características básicas, los posibles usuarios de la misma y las necesidades que éstos expresan, de las cuales se derivan las funciones de los centros de documentación de los medios periodísticos. Finalmente, se señalarán, también de forma sintética, cuáles son las actividades de los centros de documentación periodística, que desarrollan de forma práctica las funciones mencionadas.

2.2. Sobre el concepto de documentación periodística

2.2.1. Término, objeto y actividades

Sobre el **término** elegido para definir la actividad, ha existido y existe una doble opción empleada habitualmente: *documentación informativa*, por una parte, y

documentación periodística, por otra, sin perjuicio de que se hayan empleado otras denominaciones como *documentación de actualidad* o *documentación de los medios informativos*.

Sin ánimo de alargar la discusión sobre estos términos (que puede encontrarse, por citar un caso, en (FUENTES I PUJOL, 1997, p. 80-87), hemos elegido el empleo de **documentación periodística**, dado que se ajusta, a nuestro juicio, al ámbito fundamental de la disciplina: la *documentación* aplicada a los contenidos producidos con fines *periodísticos*.

Preferimos este término al de *documentación informativa*, empleado por (COLL-VINENT, 1984, p. 179) que definía además la disciplina como

“los métodos y las técnicas ordenados a centralizar las informaciones difundidas por los medios, a clasificarlas y analizarlas, y a crear los soportes adecuados para su ulterior recuperación”.

En este caso, no se tendría en cuenta una de las acepciones de *información* citadas por (LÓPEZ DE ZUAZO ALGAR, 1977, p. 106): “en sentido amplio, cualquier exposición de hechos e ideas a través de un medio de comunicación de masas”. En este concepto de “información” cabrían, pues, otros contenidos distintos a los meramente periodísticos, como la publicidad. Nos parece, por tanto, más adecuada, esta otra mención:

“(...) La documentación periodística es una subespecie de la documentación informativa.

La documentación periodística es, cabalmente, la documentación de la información periodística, como la documentación publicitaria se ocupa de la información publicitaria, contenido también vehiculado a través de los distintos medios de comunicación social”. (GALDÓN LÓPEZ, 1989, p. 31)

Una segunda variedad de acepciones radica en el **objeto** de la disciplina. López Yepes y Ros García apuntan al respecto que la documentación periodística

“puede entenderse tanto como disciplina documental al servicio de la investigación del Periodismo como al servicio de la realización de informaciones periodísticas. En este último sentido, entendida tanto actividad y proceso informativos como conjunto dependiente de la Documentación informativa, la base sobre la que se asienta es la llamada información de actualidad, siendo la noticia el objeto de la misma” (LÓPEZ YEPES y ROS GARCÍA, 1993, p. 131)

Es decir, que señala a la noticia o información como objeto de la actividad informativa, que puede ser observada como materia para el estudio científico del periodismo, por una parte; o, por otra, como objeto analizable en su uso durante el trabajo periodístico en los medios del mismo tipo.

Codina, por su parte, citado por (FUENTES I PUJOL, 1997, p. 86) apuntaba en 1994 el segundo de esos usos (los documentos utilizados para la práctica periodística) y añadía un tercero, al señalar que la documentación periodística

“designa tanto el conjunto de métodos y técnicas documentales aplicados a los documentos característicos de los centros de documentación periodística, como el objeto de tratamiento, ó sea, la información de prensa o de actualidad, en cualquier tipo de soporte”.

Ello significa que, para Codina, el objeto de la documentación periodística consistía tanto en las tareas documentales dentro del medio periodístico como en los mismos documentos empleados para ello: “la información de actualidad” o conjunto de documentos que producía el mismo medio u otros.

Sin pretender analizar en profundidad ninguna de las acepciones anteriores ni destacar la mayor validez de una u otra de ellas, el objeto tratado en el presente trabajo será el segundo de los definidos, es decir, el proceso de tratamiento de los contenidos periodísticos producidos por un medio del mismo tipo.

Otra característica del término “documentación periodística” que ha tenido, en este caso, una mayor aceptación, ha sido la noción del mismo como un conjunto global de **actividades**. No obstante, también ha habido discrepancia sobre el modo de calificar a las mismas. Como apuntábamos, Coll-Vinent las denominaba “métodos”, mientras

que Codina, en una definición posterior, cambiaba no solo el término empleado (“documentación de los medios”), sino que lo describía como

“conjunto de ciencias y técnicas documentales al servicio de: i) la producción de informaciones de actualidad, ii) el incremento de su calidad, iii) su almacenamiento y conservación, y iv) su difusión y reutilización” (CODINA, 2000, p. 49)

Sean métodos, ciencias o técnicas, en lo que sí coinciden la mayoría de los teóricos dedicados a definir la disciplina es en que el concepto agrupa una serie de actividades, aunque cada uno de ellos haga mayor hincapié en algunas de ellas. Por ejemplo, para completar la mención a las mismas, Galdón señala las siguientes:

“la valoración, selección, clasificación y archivo para su posterior uso, de textos sobre ideas, hechos, juicios y opiniones con el fin de elaborar la documentación periodística y/o difundir información documental de base periodística” (GALDÓN LÓPEZ, 1989, p.30).

Este mayor o menor énfasis es sin duda reflejo de un momento distinto, y de una concepción incluso diferente. Así, Coll-Vinent comienza mencionando la *centralización* de las informaciones, signo de que este proceso no era automático en el medio de comunicación: es decir, el centro de documentación debía localizar todas las informaciones dispersas, en muchos casos, físicamente, tanto del propio medio como de otros. Galdón hace hincapié como primera fase en la *selección* (obviando el énfasis inicial en la obtención) y en la *clasificación*, tareas que Coll-Vinent agrupa bajo el término *análisis*.

Codina es quien hace más énfasis en la tarea de mantenimiento de las informaciones analizadas (aunque no hace referencia alguna al análisis previo), especificando la necesidad de *almacenamiento y conservación*, desde una perspectiva más próxima a la recuperación de información, término utilizado en un entorno digital ajeno en buena parte a la necesidad de *creación de soportes* físicos que Coll-Vinent consideraba necesaria para la *recuperación*.

Por último, Codina y Galdón destacan expresamente la *difusión* como una tarea documental, aunque desde planteamientos teóricos diferentes. Si Galdón pone énfasis en el medio y en la forma de difusión (especificando que la información obtenida por el proceso documental se puede utilizar para elaborar nueva información periodística, o bien, para difundir directamente esa información a través de otras formas que no sean estrictamente medios de comunicación), Codina hace lo propio, pero destacando el aspecto técnico de la difusión: la posibilidad de *reutilización* de un contenido ya usado. Como puede verse, son dos concepciones de partida distintas que, en realidad, están apuntando hacia el mismo hecho: el de un contenido que deja de ser obsoleto para volver a tener un valor, periodístico o no.

Ese valor se ve, además, aumentado por el tratamiento racional de los contenidos producidos por un medio periodístico. Como apuntan (GARCÍA GUTIÉRREZ y LUCAS FERNÁNDEZ, 1987, p. 30),

“la documentación de actualidad supone, en primer término, un proceso posterior (documental) de la información de actualidad y su objetivo no es la aportación de novedad, sino la contextualización retrospectiva y testimonial de la noticia. (...) La documentación, frente al carácter espontáneo, fragmentario e imprevisible de la información de actualidad, supone un acto de racionalización, síntesis y canalización de datos retrospectivos”

Esa síntesis racional aporta nuevas capacidades a los contenidos analizados, organizados, almacenados y producidos en el proceso documental periodístico, como tendremos oportunidad de señalar a lo largo del presente trabajo.

2.2.2. Definición de documentación periodística

Una vez analizadas las definiciones anteriores, y tratando de englobar todas ellas, en lo posible, proponemos la definición de documentación periodística como el proceso por el que se realiza el tratamiento de la información periodística producida a través de un medio de comunicación, mediante un conjunto de operaciones que incluyen la valoración, selección, obtención, análisis, almacenamiento, recuperación y difusión, con el fin de que esa información sea nuevamente recibida en cualquier momento,

sea comprensible y pueda ser reutilizada por sí misma o como parte de una nueva información periodística.

En los siguientes epígrafes, ofrecemos en primer lugar una breve introducción histórica, que explica la evolución de la documentación periodística desde sus orígenes hasta la actualidad. A continuación, desarrollamos la definición ofrecida, con atención a cada uno de los elementos (características, funciones, principios y actividades documentales para el tratamiento de los contenidos) de la documentación periodística.

2.3. Etapas históricas de la documentación periodística

La documentación periodística se ha ido desarrollando en paralelo a la evolución del propio periodismo, hasta llegar a nuestros días. Se requiere, a nuestro juicio, una muestra de esa evolución que permita entender cómo han llegado a desarrollarse los elementos que definen la disciplina en la actualidad.

2.3.1. Antecedentes de la documentación periodística

La evolución de lo que puede considerarse como documentación periodística ha respondido a las distintas necesidades del periodismo por ir contando con información ya elaborada, fuera de producción propia o ajena. Esta información tenía dos finalidades fundamentales: por una parte, su posible inclusión en nuevos contenidos periodísticos; por otra parte, su utilización con fines ajenos al periodismo.

En diversas gacetas o publicaciones que relataban hechos acaecidos en distintos lugares durante los siglos XVII, XVIII y en la primera mitad del XIX se habían comenzado a realizar algunas tareas que autores como (TAPIA LÓPEZ, 1995) denominan como los antecedentes de la documentación periodística.

En estos periodos, tales antecedentes se refieren fundamentalmente, por una parte, al empleo de textos de otras publicaciones, en muchos casos extranjeras, traducidos a la lengua propia; asimismo, al primer interés por conservar ejemplares de la propia publicación; y, más adelante, siguiendo costumbres que se remontaban a la Baja Edad Media, se plantearon ciertos esfuerzos por centralizar la información publicada, por ejemplo, en un área concreta o sobre un acontecimiento específico.

Es a finales del siglo XVIII y a principios del XIX, cuando los periódicos comienzan a glosar las vidas de personas importantes recientemente fallecidas. Pronto comienzan a incorporar retratos de las mismas, que son muy costosos: de ahí que se planteara almacenarlos para su conservación y, poco a poco, para plantear su reutilización, dado que se encargaban antes del fallecimiento, en previsión de que sucediera.

Los periódicos, no obstante, debían completar esos retratos con los perfiles biográficos de las personas, previamente a su fallecimiento. Ello indujo a que los depósitos donde se almacenaran los textos, o las imágenes, o ambos a la vez, recibieran el nombre de *morgues* (SEMONCHE, 2003), por analogía con los depósitos de cadáveres.

Estas *morgues* comienzan a convertirse poco a poco en depósitos que acumulan todo tipo de imágenes, todo tipo de textos que las acompañan y que incluyen también obras de referencia con las que complementar esos textos, y con las que empezar a explicar en mayor profundidad esas imágenes.

Así ocurre en medios estadounidenses como el *Daily Advertiser*, *The New York Herald* o el *New York Tribune*, y en algún medio británico, como *The Times*, que continúan acumulando además recortes de otros periódicos, y que comienzan a usarlo para algo más que la ilustración de fotografías póstumas: también para justificar las argumentaciones políticas de los artículos de opinión.

De hecho, existe un antecedente aún anterior sobre la acumulación de recortes de periódicos: el suizo *Neue Zürcher Zeitung*, que entre 1780 y 1798 organiza las

primeras carpetas de recortes clasificados en cuatro temas: nacional, internacional, *feuilleton* (sic) y economía (GALDÓN LÓPEZ, 2002b, p. 39-40).

2.3.1.1. El nacimiento del periodismo informativo. Los primeros esbozos de documentación periodística

No obstante, es con la aparición del moderno periodismo informativo, que algunos autores como (ALBERT, 1990, p. 65) o (GARCÍA GONZÁLEZ, 1999, p. 54) sitúan en el periodo entre 1830 y 1850, cuando se datan en esta época las primeras iniciativas sobre documentación periodística. El periodismo, mayoritariamente opinativo hasta entonces, pasa a convertirse en un periodismo informativo, más enfocado a narrar hechos (noticias) que a ofrecer opiniones, y cuya pretensión es la de convertirse en un medio de comunicación de masas. Entre otros factores que contribuyen a esta transformación, pueden apuntarse:

- La aparición en Estados Unidos de la *prensa de un penique*, o dicho de otra forma, surgimiento de periódicos más baratos, con más carga informativa que opinativa, favorecidos por un aumento del número de votantes, una mejora económica general y el surgimiento de técnicas de impresión que permitían imprimir mayor cantidad de ejemplares a un menor coste (MARTIN y HANSEN, 1998, p. 24-27)
- La creación de las primeras agencias de prensa: entre otras, la agencia francesa Havas, en 1835, hoy convertida en AFP (AFP, 2003); la agencia norteamericana Associated Press, en 1848 (ASSOCIATED PRESS, 2003), la alemana Wolff (1849) y la británica Reuters (1851). Estas agencias hicieron especial uso de todas las facilidades de comunicación que se desarrollaron en estas fechas, y muy especialmente, del telégrafo (WEILL, 1994, p. 152).
- El aumento de la población urbana y alfabetizada, que podía acceder a la prensa como fuente de información, y que empleaba la prensa como una

forma de integración en la nueva sociedad (GARCÍA GONZÁLEZ, 1999, 57-58)

- El interés de algunos directores por ofrecer diarios de mayor calidad, como una forma de competir en el mercado. El paradigma de este caso es el londinense *The Times*, que aboga por ser el periódico de las clases medias ilustradas en Gran Bretaña (GARCÍA GONZÁLEZ, 1999, p. 62)

La tendencia a reducir la opinión a favor de la información comienza a hacerse palpable, como apunta (GALDÓN LÓPEZ, 2002b, p. 36-38), en Estados Unidos durante la Guerra de Sucesión (1860-1865), y a desarrollarse más aún después del conflicto (empezando por *The New York Times*, a partir de 1867), dando paso lentamente a una tendencia a explicar sobre todo los hechos más relevantes, utilizando en primer lugar las informaciones ya publicadas por el propio periódico.

Comienzan, pues, al menos los grandes medios a elaborar bibliotecas de referencia: como apunta (SEMONCHE, 2003), ya en 1845, al menos en el *New York Herald*, y más adelante, índices. En este caso, aunque el primer índice conocido era el *Palmer's Index to The Times*, en Gran Bretaña, que se venía publicando desde 1790, fue solo a partir de mediados del XIX, empezando con *The New York Times*, desde 1851, cuando algunos grandes medios comenzaron sistemáticamente a producir índices de sus publicaciones.

Es también en este periodo cuando los archivos de recortes, principalmente biográficos comienzan a implantarse de modo más habitual en algunos grandes periódicos estadounidense, como el citado *The New York Times*, en 1869, o *The New York World*; este último, aunque comenzó a recortar materiales en 1870, solo lo organizó como un verdadero departamento de referencia en 1889. Otros periódicos, como el *Boston Globe*, el *Springfield Republican* o el *San Francisco Chronicle*, también organizaron a finales de siglo una biblioteca de referencia y un archivo de recortes.

En Europa, aunque tímidamente, también comenzaron a surgir iniciativas que dan lugar a que se pueda hablar del uso de la documentación en las redacciones de los

periódicos. Dos casos concretos son los del citado *The Times*, cuya sección de Internacional acumula desde 1883 todo tipo de material de referencia; y, más aisladamente, el también británico *The Guardian*, cuyo corresponsal en Egipto, Evans, ofrece una actualización diaria de los mapas durante una de las guerras sufridas en ese territorio en la década de los 80.

No obstante, estas y otras iniciativas aisladas no formaban aún parte de un centro de documentación periodística. Cada uno de estos servicios (biblioteca de referencia, archivo de recortes, así como archivo fotográfico) estaban situados físicamente en diferentes lugares de la redacción y, más aún, dependían de diferentes secciones del medio. Además, en muchos casos, estas recopilaciones cesaban, por falta de medios, y eran luego cedidas a distintas bibliotecas o, simplemente, desaparecían.

Por otra parte, muchas pequeñas empresas periodísticas no disponían de medios, ni materiales ni técnicos ni humanos, para poder implantar iniciativas documentales. Los primeros balbuceos documentales eran aptos solo para grandes medios.

2.3.2. Los primeros centros de documentación periodística

(GALDÓN LÓPEZ, 2002b, p. 45-47) apunta las primeras dos décadas del siglo XX como el periodo en el que las antiguas *morgues* se convierten ya en departamentos de documentación más o menos institucionalizados. El primero caso sería el del diario británico *Daily Mail*, donde en 1896 ya se realiza una clasificación de recortes por materias y personas, tanto del propio periódico como de otros británicos y extranjeros, disponiendo además de un índice de contenidos y de una biblioteca de referencia. Algo parecido sucede en *The Times*, aunque solo en 1909 se institucionaliza un servicio de documentación como tal.

En Estados Unidos, muchos periódicos comienzan a disponer de biografías compuestas no solo para ilustrar el fallecimiento de personas, sino para ofrecer información sobre personajes vivos. A ello se unen prácticas como los sistemas de índices de referencias cruzadas y de índices de tarjeta en algunos medios, así como la

catalogación de las obras de referencia disponibles, la revisión y elaboración en diversos medios de archivos de recortes de múltiples periódicos, tanto locales como incluso extranjeros; y, sobre todo, la dedicación de varias personas al incipiente centro de documentación. Algunos periódicos de habla hispana siguen esta senda, como *La Nación*, de Argentina, en 1910, o el español *El Debate*, en 1911.

La Primera Guerra Mundial provocó un notable aumento del uso de la documentación en la prensa. *The Times*, por ejemplo, utiliza todo el citado material acumulado por la sección de Internacional para realizar una “Historia de la guerra” quincenal. En *The New York Times*, al aumentar la colección de recortes para cubrir la guerra, decidieron realizar encabezamientos de materia para tratar de organizar los recortes. (SEMONCHE, 2003).

2.3.2.1. La toma de conciencia: el desarrollo en los años 20.

El final de la guerra, y el seguimiento de sus consecuencias, supone un aumento de la necesidad de información complementaria a la mera narración de los hechos. El primer caso concreto sería la elaboración por parte de *The New York Times* de información continua, con la inserción de biografías o de datos sobre los ejércitos de los países contendientes, en la Conferencia de Versalles que pone fin a la entonces denominada Gran Guerra.

Los años siguientes suponen la toma de conciencia definitiva sobre la importancia de contar con un buen “servicio de referencia”, al igual que en otras disciplinas se concebía la importancia de la documentación especializada: los grandes medios, al menos en Estados Unidos, comenzaron a crear sus propios centros de documentación.

Al mismo tiempo, comienza a tomarse conciencia de la importancia de un buen centro de documentación periodística, también a nivel teórico. La asociación estadounidense dedicada a los centros de documentación especializados, la Special Libraries Association (SLA), creó en 1913 un comité para analizar el manejo de los

archivos de recortes de artículos, y en 1923 una sección específica sobre las “Newspaper Libraries” (los centros de documentación en periódicos), que se mantiene en la actualidad con el nombre de “News Division” (SEMONCHE, 2003).

En esta década, se plantea la centralización del servicio de documentación (GONZÁLEZ QUESADA, 1995, p. 25-26), para evitar la duplicidad que en muchos medios existía al tener separadas la biblioteca de referencia, por una parte, y el archivo de recortes, por otra, con el fin de ahorrar en costes (los de mantener dos servicios por separado), en rapidez (al poder acceder simultáneamente a ambos servicios) y en flexibilidad (era más fácil atender consultas más complicadas al tener todas las herramientas documentales en un mismo entorno).

También se desarrolló, aunque solo sea en los grandes medios, la difusión de los productos documentales al público en general. Es el caso de *The Times*, que contenía además en el índice respectivo las referencias a los artículos, incluidos los suplementos del periódico y los contenidos del dominical *The Sunday Times*. Por su parte, *The New York Times* comenzó a comercializar sus índices en 1930; su aceptación fue tal que se convirtió en obra habitual en bibliotecas y centros educativos de Estados Unidos, junto con la colección completa del diario.

2.3.3. El trabajo documental en prensa: 1920-1945

2.3.3.1. El periodismo interpretativo

El aumento en el uso de las explicaciones complementarias a la narración informativa de los hechos evoluciona hacia lo que se denominará el **periodismo interpretativo**: tratar de explicar en mayor profundidad sobre todo los hechos más importantes, que se van contando a diario pero que necesitan un análisis, una recopilación de acontecimientos y una contextualización. Esta necesidad se ve acuciada ante el surgimiento de un mundo cada vez más complejo, sensación aumentada por la crisis económica que surgió en 1929.

Dos son las fórmulas que surgen, en el periodismo estadounidense (GALDÓN LÓPEZ, 2002b, p. 77-87), para dar paso a este tipo de información. La primera de ellas es el llamado *newsmagazine*, o revista semanal de noticias, cuyo primer caso fue *Time*, en 1923. En esta y en otras revistas, comienzan desde el principio a tener un equipo de personas, denominadas “verificadoras” o “investigadoras”, que se dedican a cotejar unos datos con otros, comprobarlos e interrelacionarlos.

El segundo modelo de periodismo interpretativo lo constituye el *dominical*, cuyo primer exponente se publicó en *The New York Times*. El dominical se basa en ofrecer documentación sobre noticias seleccionadas, y establecer una visión completa de conjunto, con el objetivo de situar cada hecho seleccionado en su contexto, suministrando elementos que aporten significado. Muchos de esos elementos tienen su origen en el servicio de documentación, que comienza por tanto a ser potenciado, en recursos materiales, en cantidad de personal y en categoría de las mismas.

El caso de *The New York Times*, empero, no estaba extendido, y menos, fuera de los Estados Unidos. El propio dominical de este periódico necesitará década y media hasta que, en 1945, el periodismo interpretativo, y con él, el uso de la documentación, sea considerado como un elemento reconocido por todos los medios. En ese reconocimiento puede haber influido notablemente la Segunda Guerra Mundial, que provoca el desarrollo de la interpretación de noticias.

2.3.3.2. Aumento de los servicios de documentación periodística. Primeras concepciones teóricas

Fuera de Estados Unidos, la documentación periodística entre 1920 y 1945 aún avanzaba a pasos muy lentos en general. Solo algunos periódicos establecen ya servicios de documentación, conscientes de su valor, especialmente, en Gran Bretaña (*The Times*, *The Guardian*, *Yorkshire Post*...). También surgen en esta época los primeros centros de documentación de algunos periódicos como el barcelonés *La Vanguardia* (1920, remodelándolo en 1939), el noruego *Aftenposten* (1928), el holandés *De Telegraaf* (1929), el belga *Gazet van Antwerpen* (1930)...

A comienzos de la década de los 30, surgieron ya los primeros estudios de rango universitario sobre documentación periodística en Estados Unidos, así como los primeros manuales sobre la materia. Entre ellos, el de R.W. Desmond, *Newspaper Reference Methods*, en 1933 (GALDÓN LÓPEZ, 2002b, p. 62-71). Desmond ya citaba, entre otras funciones informativas del centro de documentación de prensa, las siguientes:

- Evitar errores
- Suministrar antecedentes de la noticia
- Mejorar la presentación de la información local
- Preparar material por adelantado y sugerir informaciones
- Servir a los editorialistas
- Realizar campañas en beneficio de la comunidad
- Hacer asequible el material especializado
- Conservar la documentación valiosa
- Dar buena imagen del periódico
- Servir al departamento de publicidad
- Servir al departamento de difusión.

Como acertadamente apunta (GONZÁLEZ QUESADA, 1995, p. 26), en esta división subyacía una segunda distinción que separaba dichas funciones entre aquellas que tenían una proyección directa sobre los lectores (las siete primeras) y las funciones vinculadas a la economía de la propia economía periodística (las relativas al archivo de documentos valiosos, la buena imagen del medio y los servicios a los departamentos de publicidad y difusión). Desmond ya argumentaba entonces que el valor de un archivo no estaba en el número de recortes, sino en la habilidad con la

que éstos habían sido seleccionados y analizados, con el fin de ahorrar espacio y agilizar tanto la consulta como la recuperación de la información.

Tales funciones se formalizaban en la práctica en una serie de actividades concretas realizadas, según (GALDÓN LÓPEZ, 2002b, p. 63-70) , en buena parte de los periódicos de Estados Unidos:

- La citada centralización de los trabajos documentales; en un solo departamento, fusionando el archivo de recortes y la biblioteca de referencia.
- La elaboración de índices del propio periódico.
- La selección de recortes, considerada más bien como una eliminación de aquellos recortes cuya información ya estuviera contenida en obras de referencia poseídas por el propio medio.
- La clasificación de los materiales, que era considerada como la tarea más importante. Cada medio aplicaba su propio sistema de clasificación a los recortes periodísticos de que disponía, agrupándolos en carpetas básicamente por tema, por personajes y por país; secundariamente, otros medios también agrupaban por fecha. En cada medio se trataban de fijar encabezamientos y subdivisiones simples, que fueran fáciles de recordar. El sistema, asimismo, solía permitir la codificación de un mismo artículo por varias entradas.
- El archivo: para ello, se emplearon, en los medios más avanzados, equipos de archivo específicos para cada material de documentación (fotografías en diversos soportes, recortes de papel, etc.)
- La recuperación de los materiales a petición de los periodistas. El objetivo, ya entonces, era servir la información en el menor tiempo posible: un grado de respuesta positiva del 80% era aceptable, según Desmond (citado por (SEMONCHE, 2003). Las peticiones se recibían de todas las secciones del

periódico. También se recibían, aunque en menor medida, peticiones de usuarios externos al periódico, especialmente, utilizando los archivos con valor histórico.

2.3.4. El desarrollo documental a partir de 1945

Desde 1945, se produce un aumento de los medios periodísticos que incorporan un centro estable de documentación o que lo mejoran, en su caso. En los diez años siguientes, la mayoría de los grandes periódicos estadounidenses así lo hacen, al igual que en otros países. Es el caso de Gran Bretaña, donde aumentan enormemente los servicios de documentación en los medios ya existentes (*The Times*, *The Guardian*) o surgen otros nuevos medios que incorporan ya ese servicio (*Financial Times*, *Daily Mirror*).

También en Alemania y Japón, la documentación periodística alcanza, según (GALDÓN LÓPEZ, 2002b, p. 96), en 1970 cotas comparables en algunos de sus medios a las de los periódicos estadounidenses. En Alemania, en particular, se parte de una fuerte tradición biblioteconómica en sus instituciones públicas que se amplifica entre 1945 y 1970 en los medios periodísticos germanos, hasta el punto de que en 1970, todos los periódicos con una tirada de más de 100.000 ejemplares, y el 75% de los periódicos con menos de 100.000 ejemplares, disponían de un servicio de documentación. En otro ámbito, destaca la revista de noticias *Der Spiegel*, de periodicidad semanal, surgida en 1947, y cuyo servicio documental en 1965 ofrece un nivel comparable al de sus equivalentes en Estados Unidos.

Esta evolución es menor en otros países, aunque también aumenta el número de medios que le dan importancia a su servicio de documentación. Es el caso, en Francia, del diario *Le Monde* o del semanario *L'Express*. En España, según (GALDÓN LÓPEZ, 2002b, p. 97), solo *La Vanguardia* y, a partir de 1948, *ABC* (AGUADO, 1995), disponen de servicio de documentación.

Tal desarrollo se vio aumentado por la llegada de un nuevo medio de comunicación: la televisión. El desarrollo de este medio, y la difusión de informaciones periodísticas

a través de él, influyó en el uso de la documentación periodística. Por una parte, como apunta (GONZÁLEZ QUESADA, 1995, p. 28), la capacidad de la televisión para informar de manera inmediata hizo que la prensa tuviera que ofrecer esa información desde otro ángulo, contando “lo que subyacía bajo los hechos”, para lo cual se valía del uso masivo de la documentación.

(SEMONCHE, 2003) apostilla a esta circunstancia que, además, fueron las cadenas televisivas las que comenzaron a emplear los contenidos de la prensa y de la radio con fines documentales, fomentando paradójicamente el uso de fondos documentales tanto en los medios de prensa escrita como en los audiovisuales, que también comenzaron a organizar sus fondos producidos expresamente (material escrito, pero también sus fondos sonoros y audiovisuales).

2.3.4.1. Problemas técnicos y primeras soluciones

El aumento de los servicios de documentación periodística y, especialmente, de su tamaño, también trajo como consecuencia la existencia de una gran cantidad de material que provocaba enormes problemas:

- De clasificación: un mayor número de documentos supone la necesidad de dedicar mayor esfuerzo a esta tarea.
- De custodia: el número cada vez mayor de documentos supone que sea mucho más difícil mantener físicamente todos ellos. Surgen problemas de espacio físico, de posible pérdida de los mismos por resultar más difícil su control, etc. La saturación de los fondos documentales se convierte en el principal problema de los centros documentales, y el expurgo de aquellos documentos considerados como no útiles, en una de las principales tareas.
- De recuperación: al ser mayor el número de documentos disponibles, también resulta más difícil encontrar aquellos documentos relevantes para una consulta dada con las herramientas documentales tradicionales. Además, como apuntan (HERNÁNDEZ PÉREZ y CARIDAD

SEBASTIÁN, 2000, p. 60), al emplear el sistema de archivo por carpetas, una misma noticia puede estar en más de una carpeta (por ejemplo, en una carpeta sobre un personaje concreto que participe en la noticia, en otra carpeta sobre el tema de la noticia, etc.), de modo que para consultas sobre temas genéricos, era habitual buscar en más de una carpeta, con la consiguiente ralentización.

Estos problemas, compartidos por los documentalistas especializados en otras áreas, provocaron la búsqueda de nuevas soluciones, que pueden organizarse de menor a mayor coste técnico y económico:

- La creación de nuevos métodos de archivo, que reducían espacio al tiempo que mejoraban la localización centralizada de los documentos.
- El desarrollo de las microfilmaciones de colecciones completas, iniciadas en Estados Unidos, en el *New York Herald Tribune*, a finales de los años 20 (SEMONCHE, 2003). Esta técnica llegó a ser generalizada en Estados Unidos ((GONZÁLEZ QUESADA, 1995, p. 28) apunta que el 90% de sus periódicos tenía microfilmados sus ejemplares en 1970), así como en Japón, aunque apenas se aplicó en Europa hasta esa fecha.

Asimismo, un número mucho menor de medios también microfilmó sus archivos de recortes: apenas el 25% de los periódicos estadounidenses y, en Europa, un puñado de medios, como el citado *Der Spiegel* o el diario suizo *Neue Zürcher Zeitung*. (GALDÓN LÓPEZ, 2002b, p. 115).

- Los índices asociados con la microfilmación de periódicos. Según la cronología de (FOLEY y BRISCOE, 1989, p. 18), la primera iniciativa en este sentido fue en 1964, en el diario estadounidense *Rochester Times-Union and Democrat & Chronicle* (sic). Esta solución pretendía reducir la necesidad de que el propio medio hiciera recortes con sus propios artículos. Fue una iniciativa que, en 1970, apenas era seguida por una cincuentena de periódicos en todo el mundo, teniendo como principales impulsores a *The*

Times, The New York Times y The Guardian (GALDÓN LÓPEZ, 2002b, p. 117).

- Las archivadoras electromecánicas, que sustituían a las estanterías tradicionales y estaban adaptadas a requerimientos concretos de cada medio. Por su coste, también fueron usadas por muy pocos medios.

2.3.5. Aplicación de la informática a la documentación periodística

Aunque las primeras iniciativas de informatización en servicios documentales de prensa datan de finales de los 50 (FUENTES y CONESA, 1994, p. 43), fue a finales de los 60 y comienzos de los 70, cuando se planteó el uso de los ordenadores aplicados a la documentación periodística como algo más que un proyecto. En 1967, tras un seminario sobre documentación periodística organizado por la asociación de editores de periódicos norteamericanos, ANPA, se estableció un comité específico para la aplicación de sistemas informáticos a los centros de documentación periodística en Estados Unidos (FOLEY y BRISCOE, 1989, p. 18).

Desde finales de los años 70 y, especialmente, durante toda la década de los 80, como señalan (FUENTES y CONESA, 1994, p. 44-45) se desarrollaron las primeras bases de datos periodísticas, como parte de una corriente en la que, además de disponer de mayor capacidad de memoria, los ordenadores eran ya instrumentos de trabajo que se fueron convirtiendo en habituales, al menos, en las redacciones de los grandes periódicos de todo el mundo.

En la década de los 70, surgió el debate en torno a una doble posibilidad:

- La utilización de sistemas informáticos que permitieran recuperar la referencia sobre la información almacenada, manteniendo el almacenamiento en microfilms o en microfichas de forma más o menos inmediata. Un primer ejemplo: el *Detroit Free Press* presentó en 1971 su

sistema informático Miracode, que permitía manejar el microfilmado de recortes de que disponía.

Un sistema similar fue el de *The New York Times*, que ya en 1972 comenzó a hacer disponible su servicio New York Times Information Bank (CARIDAD, 1984, p. 145-165), una base de datos de resúmenes de sus artículos accesible en línea a través de un terminal y una conexión telefónica, utilizando además tesauros como modo de acceso. Este servicio estaba más orientado a usuarios externos que a los propios redactores del periódico.

- El empleo de sistemas informáticos que almacenaran sus contenidos directamente en soportes magnéticos (cintas y, posteriormente, discos), y no en microfilms, permitiendo recuperar inicialmente términos obtenidos en los resúmenes hechos sobre el tema; y, más adelante, posibilitando las búsquedas en el texto completo (FOLEY y BRISCOE, 1989, p. 18-20).

Durante la década de los 70, la solución que compaginaba el uso de los índices en ordenadores con la recuperación final en microfilm o microficha fue la elegida en principio por los medios que se animaron a realizar la correspondiente inversión. Diversos sistemas similares se comercializaron en varios periódicos estadounidenses y europeos en los 10 años siguientes. Incluso en España, *El Correo Español-El Pueblo Vasco*, de Bilbao, comenzó a experimentar con un sistema similar más tarde, en 1983.

Pero esa opción tecnológica presentaba dos grandes problemas: uno, su carestía, que le hacía solo accesible para grandes medios dispuestos a realizar una costosa inversión sin tener clara su rentabilidad a medio plazo. Y dos, que seguía sin solucionar los problemas de recuperación.

Ello hizo que, de modo experimental durante los años 70, y de un modo ya más desarrollado, técnica y comercialmente, las bases de datos cuyos contenidos estaban almacenados en soportes magnéticos, directamente recuperables desde terminales, empezaran a imponerse a comienzos de los años 80.

2.3.5.1. Las bases de datos

Aunque estas bases de datos eran empleadas inicialmente dentro de los propios medios, pronto comenzaron a ser implementadas para su difusión en línea a través de redes telemáticas, accesibles por clientes externos.

Fue a finales de los 70 cuando surgieron las primeras bases de datos a texto completo. Estas últimas fueron desarrolladas, inicialmente, para uso interno de algunos medios (el citado *Boston Globe*, el *St. Louis Post Dispatch* o el *Philadelphia Inquirer*, en Estados Unidos; *The Globe & Mail*, en Toronto, Canadá, etc.), aunque pronto se comenzó a extender la idea de crear bases de datos que contuvieran simultáneamente la información publicada por múltiples medios para su difusión en línea. En algunos casos concretos, los medios formaban parte del mismo grupo de prensa, como el norteamericano *Knight-Ridder*, que poseía en 1982 una treintena de periódicos interconectados mediante tres bases de datos centrales de uso interno.

Pronto, sin embargo, comenzaron a desarrollarse empresas independientes de distribución de bases de datos como elemento tecnológico que ofrecían, mediante acuerdos con los medios, las referencias sobre los artículos y, en algunos casos, los contenidos textuales completos de los mismos a cualquier cliente que pudiera pagarlo. Es el caso de Nexis, empresa norteamericana que ofrecía a comienzos de los 80 los contenidos de medios como *The Washington Post*, *Newsweek*, *U.S. News & World Report*, etc. (PROVENZANO, 1981).

(FUENTES y CONESA, 1994, p. 43) señalan que el procedimiento, en cualquiera de las dos opciones (el propio medio mantenía su base de datos o la cedía a un distribuidor) consistía en traspasar los textos de las cintas magnéticas usadas por el periódico, en el proceso de fotocomposición, a las bases de datos, previa limpieza de los códigos de fotocomposición y de la información que no iba a ser incluida en la base de datos (anuncios, espectáculos, agendas, predicción del tiempo, etc.).

A estas bases de datos se les aplicaban técnicas de indexación automática, que permitían posteriormente buscar por cualquier palabra existente en cualquier artículo, y se podía recuperar automáticamente el texto completo de los artículos obtenidos

tras la búsqueda. Algunos medios, como *Le Monde*, complementaban esa indexación automática con una indización manual, mediante palabras clave o descriptores de un tesoro.

2.3.5.1.1. Inconvenientes de las bases de datos

El empleo de las bases de datos, aunque aportaba grandes ventajas al centro de documentación, presentaba también algunos problemas.

El primero de ellos, al menos, durante los años 80 y 90, fue que se perdía toda la información icónica accesible en el medio en papel o en microfilm: no solo las fotografías e ilustraciones, sino otros aspectos como la tipografía usada o la distribución dentro de cada página y de cada ejemplar del medio.

Por otra parte, muchas de las bases de datos, al menos durante los 80, y en particular, las bases elaboradas por los distribuidores, y no por los medios, eran fundamentalmente de resúmenes, lo cual causaba desconfianza en su uso por parte de los periodistas de los medios que estaban suscritos a esas bases.

Un problema adicional era la complejidad del manejo de estas bases, que hacía difícil su consulta si no se conocía con exactitud qué posibles bases existían, cuáles eran sus fondos y cómo consultarlas. En este sentido, (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 63) recogen la opinión de Nora Paul sobre el cambio de papel de los documentalistas debido al uso de las bases de datos: pasaron de ser conservadores de documentos a convertirse en intermediarios de los mismos entre las fuentes originales y los periodistas.

Otro problema relacionado con el anterior, especialmente durante los años 80, era el enorme coste del acceso a una base de datos en línea: no solo se cobraba por la información, sino por el tiempo de acceso a la misma. La complejidad en la búsqueda, unida al precio de la consulta, provocaba que ésta fuera prohibitiva para muchos medios, reduciendo este recurso en muchos casos solo a aquellos grandes medios que podían permitirse su uso.

Por otra parte, el uso de las bases de datos no podía dar una respuesta precisa a búsquedas que precisaban una visión general de un tema amplio, y tampoco se podían consultar varios medios simultáneamente, salvo que estuviesen integrados en la misma base de datos. Éste último detalle hizo que los grandes distribuidores de bases de datos, como Nexis (convertida más tarde en Lexis-Nexis) o Profile trataran de incluir el mayor número posible de medios entre sus contenidos, compartiendo con cada medio los beneficios obtenidos por la consulta de los contenidos del mismo, pero cobrándole a éste, a cambio, la consulta a contenidos de otros medios.

Además, la mayoría de estas bases de datos contenían fondos de medios anglosajones, y en inglés. Por citar el caso español, solo dos distribuidores españoles ofrecían a comienzos de los 90 contenidos de medios españoles: Baratz, desde comienzos de los 80, y Documentación de Medios (FUENTES I PUJOL, 1994, p. 23).

Caso aparte son las grandes agencias de prensa (AP, Reuters, Efe..., pero no solo), que se convierten en las grandes productoras de bases de datos de texto completo, aunque su uso presente como gran inconveniente la redundancia: en los fondos de estas bases de datos se incluyen todos los despachos de agencia, lo que incluye todos los que se publican sobre un tema concreto, por ejemplo, en el mismo día, con apenas variaciones entre los mismos.

2.3.5.2. Memorias ópticas y GED

La década de los 90 supuso una gran transformación, al menos desde el punto de vista tecnológico, del periodismo y, con él, de la documentación periodística. El primero de los desarrollos de esta época fue el uso de los CD-ROM, que, si bien existían como soporte desde comienzos de los 80 para la grabación de música, pasaron a usarse a comienzos de los 90 para recopilar todas las publicaciones de un medio, tanto en su versión textual como en su versión icónica (es decir, tal y como se visualizaba cada página del medio en papel, pero visible en la pantalla del ordenador).

En este caso, además de algunos ejemplos anglosajones (como los diarios *The Times* o *The Guardian*), existieron notables ejemplos también en España, como *El Mundo*, *La Vanguardia* o *El Periódico*, hacia mediados de la década de los 90.

Si bien poseían una notable capacidad, los CD-ROM presentaban un serio problema como soporte documental: recopilaban noticias, por ejemplo, de un semestre... después de que sucediera. Ello suponía que debía esperarse durante un lapso concreto para obtener almacenadas todas las noticias de un periodo, con lo cual el CD-ROM tenía un valor histórico, sociológico o incluso periodístico, pero siempre a medio plazo.

El aumento de las capacidades de almacenamiento informático, y la generalización de las memorias ópticas y, en particular, de los CD-ROM, permitió que se plantearan alternativas como la denominada “gestión electrónica de documentos” (GED), una de cuyas aplicaciones consistía en crear bases de datos que contuvieran tanto el texto completo como reproducciones facsímiles de cada una de las páginas de las publicaciones almacenadas, así como referencias documentales sobre cada artículo las mismas. Esta opción, que permitía plantear una alternativa al uso de microfilms, fue desarrollada, aunque lentamente, dada su carestía, por algunos grandes medios de prensa, como el grupo editorial alemán Gruner & Jahr (FUENTES y CONESA, 1994, p. 46-47).

2.3.6. El desarrollo de los centros de documentación en España

El desarrollo mayoritario de los centros de documentación en medios periodísticos no siempre reflejó estos avances tecnológicos. En los 70 y, en buena medida, en los 80, muchos de los grandes o medianos periódicos del mundo estaban al margen de estas discusiones tecnológicas. Los diversos estudios recogidos por (GALDÓN LÓPEZ, 2002b, p. 132-143) en diferentes países europeos y españoles señalaban en los 70 una situación que establecía un retraso de casi 50 años en los diferentes servicios de documentación de periódicos europeos: en los países escandinavos, en

su mayoría, se trataban de archivos simples de recortes y, en algunos casos, de tarjetas indizadas.

En España, a principios de los 70, medios como el diario barcelonés *El Noticiero Universal* desarrollaron desde 1972 iniciativas documentales, ya sean la elaboración de un índice de contenidos, el recorte y clasificación de artículos, o la microfilmación de la colección completa del diario, nacido en 1888. Otros medios como *Pueblo* o *El Correo de Andalucía* comenzaron a mantener un servicio de documentación en estas fechas. Y especialmente, el nuevo diario *El País*, nacido en 1976, incorporó ya un servicio de documentación que seguía en parte las pautas mantenidas por el diario francés *Le Monde*.

Esta casi absoluta sequía en la documentación periodística española, se debe, según (COLL-VINENT, 1978, p. 25-26), a la falta de libertad informativa durante el régimen franquista, que provocó situaciones como la carencia, en 1978, de servicios de documentación en pequeños diarios españoles, o que los grandes diarios en Madrid y Barcelona dispusieran de pequeños archivos simples, y entre una y cinco personas dedicados a él, detalle éste último confirmado en 1984 por (FUENTES I PUJOL, 1984, p. 101). Las empresas periodísticas, tal vez por el coste que pueda suponer y por la falta de rentabilidad a primera vista, se resistían por entonces a realizar mejoras de cualquier tipo en sus medios.

Durante la década de los 80, y con las excepciones de algunos de los medios más destacados, según indica (GONZÁLEZ QUESADA, 1995, p. 32), “la tendencia a la automatización y el desarrollo de los servicios documentales es lenta, y muchos periodistas todavía recurren a sus archivos personales”.

A pesar de ello, algunos periódicos españoles comienzan a mejorar, aún parcialmente, los medios de los que disponen. Es el caso de nuevos periódicos, como el vasco *Deia* o de otros medios con décadas de historia, como el zaragozano *Heraldo de Aragón*, que no solo aumentan sus medios, sino que asumen la necesidad de contar con gente que posea una formación adecuada. A finales de los 70 y comienzos de los 80, se comienzan además a difundir comercialmente, por parte de

los dos periódicos españoles de mayor tirada, *El País* (desde 1976) y *ABC* (desde 1983), algunos de sus productos documentales: los índices y las microfichas.

El número de centros de documentación aumentó notablemente, y comenzaron a plantearse iniciativas para la automatización de algunos de esos centros, aunque lentamente.

De hecho, como puede observarse en recopilaciones como el seminario celebrado en 1992 en Valencia (BELLVESER, 1994), al que acudieron una treintena de medios de toda España, en esa fecha, buena parte de los medios de todo tipo (desde grandes diarios a semanarios de información general o revistas especializadas, así como los medios audiovisuales) habían comenzado a generar bases de datos propias, si bien eran en muchos casos referenciales, recuperando los contenidos completos en microfilm o, preferentemente, mediante los clásicos archivos de recortes, con la excepción del diario *El Sol*, modelo en muchos aspectos técnicos (y especialmente en su pretensión de obtener una “redacción sin papeles”), pero de corta vida.

No obstante, en otros muchos medios pequeños, según diversos estudios citados por (DELGADO SOLER, 1999) sobre el estado de los centros de documentación en las comunidades de Andalucía, Murcia, Valencia y Madrid, se carecía no ya de grados de automatización, sino de mera organización del centro documental.

2.3.7. El comienzo de la difusión periodística a través de Internet

A comienzos de la década de los 90, la web comenzó a ser desarrollada, tímidamente, a comienzos de los 90, como plataforma de difusión de contenidos académicos. Muy pronto, ya en 1994, los primeros medios periodísticos, que habían ensayado con relativo éxito otras plataformas de difusión en línea, empezaron a colocar sus primeros contenidos en páginas y sitios web.

El *Daily Telegraph*, en una versión solo texto, y el *San José Mercury News*, en su versión íntegra, fueron los primeros ejemplos de periódicos accesibles en Internet, en

1994 (ARMENTIA, CAMINOS y otros, 2000, p. 58), en un ejemplo pronto seguido por periódicos de todo el mundo ante la facilidad y sobre todo, el relativamente bajo coste de esta modalidad de difusión. De hecho, las primeras publicaciones españolas en Internet datan de finales de 1994 (la revista valenciana *El Temps*) y de comienzos de 1995 (*La Vanguardia*, *El Periódico*, *Avui* y *ABC*), a los que siguieron, en 1996, *El Mundo* y *El País*.

Tras cierta incertidumbre inicial, los medios comenzaron a aprovechar el hecho de disponer de los contenidos en formato digital para ofrecer el acceso retrospectivo a los documentos ya publicados en lo que se ha convenido en denominar **servicios de valor añadido** (FUENTES I PUJOL y GONZÁLEZ QUESADA, 1998). Pronto, algunos medios observaron la posibilidad de rentabilizar el acceso a esos contenidos, del mismo modo que se obtenía ingresos por la venta de ejemplares atrasados. De esta forma, a partir de 2000, el acceso a esos fondos retrospectivos y, sobre todo, a la hemeroteca, comenzó a ser en algunos casos de pago.

Más adelante, algunos medios comenzaron a plantear la posibilidad de simultanear el acceso a la hemeroteca, o a documentos o informes concretos, mediante pago, con el acceso concreto a los documentos publicados recientemente, mediante libre acceso. Los medios comenzaron a ofrecer el acceso a esa información retrospectiva como argumento de calidad.

No obstante, en general, esta información retrospectiva suele carecer de tratamiento documental, al menos, en los medios de comunicación españoles accesibles a través de la web. En muchos casos, este acceso consiste en la mera implantación de un buscador que, en muchos casos, no permite recuperar adecuadamente la información deseada.

(JIMÉNEZ LÓPEZ, 2001, p. 360-363), en un estudio monográfico al respecto, señala algunos motivos: la falta de documentalistas para poner en marcha estos servicios; la falta de planificación en su elaboración; la mala estructuración de estos servicios, que incluye la dispersión del acceso de los mismos entre las diversas páginas de un mismo sitio web del medio; la falta de tratamiento documental sobre

los contenidos; por último, el hecho de no evaluar el funcionamiento de estos servicios por parte del medio.

2.4. Elementos documentales de la información periodística

Los rasgos claves que definen la documentación periodística, han sido estudiados por diversos autores, entre los que cabe destacar a García Gutiérrez y Lucas Fernández, a Fuentes y Conesa, y a Galdón.

No obstante, al comprobar las opiniones de estos autores y de otros que se refieren a ellos, como López Yepes, se constata de nuevo que se combinan en algunos casos los rasgos o principios que definen la documentación periodística, entendida como proceso, con las características de los documentos periodísticos, objetos de ese proceso.

Por otra parte, ese proceso debe tener como punto de partida las necesidades de los destinatarios finales de ese proceso: principalmente, los periodistas, pero no solamente. De hecho, el uso de los productos periodísticos en línea, como hemos indicado, puede potenciar a otro tipo de usuarios que, si bien no son nuevos en el uso de la información periodística entendida como producto de un proceso documental, sí han usado en mucha menor medida esta documentación, al tener mayores restricciones sobre el acceso a los fondos del centro de documentación.

Definiremos, pues, las características genéricas de los documentos periodísticos; las necesidades de información expresadas, al menos, por sus principales usuarios, los periodistas; finalmente, los principios del proceso de la documentación periodística, que pretenden responder a las características de los documentos y a las necesidades de los usuarios.

2.4.1. Características de los documentos periodísticos

(GARCÍA GUTIÉRREZ y LUCAS FERNÁNDEZ, 1987, p. 19-27) definieron una lista, mantenida más tarde por (GARCÍA GUTIÉRREZ, 1999b, p. 214-216), de características aplicables a los documentos periodísticos, que seguiremos en su mayoría, aunque en distinto orden y complementándolas con aportaciones propias, o bien, tomadas de otros autores:

- **Enciclopedismo:** un medio de prensa de información general trata, por definición, cualquier tipo de tema, aunque ciertamente haya unos temas a los que dedique mayor cantidad de espacio y de tiempo. Esta multiplicidad obliga al centro a adoptar una serie de criterios que permitan agrupar esos temas, por ejemplo, en categorías generales, de forma que la relación entre contenidos publicados en diferentes momentos sea un criterio para que esos contenidos puedan ser recuperados conjuntamente.

El hecho de poder abarcar cualquier campo temático complica el análisis documental de contenido, evitando, por ejemplo, que puedan aplicarse técnicas documentales empleadas con éxito en áreas temáticas muy concretas.

Asimismo, un mismo campo temático, ya sea concreto o específico, puede estar nombrado de diferentes formas (por ejemplo, puede haber diferentes referencias a un mismo personaje en noticias relacionadas con el mismo), lo cual requiere que el centro emplee criterios para normalizar todas esas referencias, para facilitar la recuperación de todos los documentos sobre un tema concreto.

- **Universalidad.** Entendida en estos sentidos:
 - *Procedencia geográfica:* un documento periodístico puede proceder de cualquier punto del mundo.

- *Fuente*: el autor inicial de los contenidos que luego se refieran en un documento periodístico puede ser cualquier persona física o jurídica; ya sea una declaración personal de un individuo en una entrevista o una obra de múltiples autores, o una declaración asumida por una institución o un colectivo, por citar algunos casos.
- *Canales de transmisión*: los documentos periodísticos pueden ser difundidos y recibidos a través de múltiples medios: prensa, radio o televisión, además de otros múltiples soportes hoy disponibles para la difusión de los medios digitales.
- *Audiencias*: los documentos periodísticos pueden ser difundidos a todo tipo de públicos, en momentos muy determinados, lo cual supone que en su composición se den por consabidos (y por tanto, no se expliciten) detalles contextuales necesarios para la comprensión de la información o de la opinión que se destacan, dándose por hecho que, en el momento de la difusión, esos detalles son conocidos por todos los potenciales destinatarios de la información. Esta presuposición implica un grave desafío para el análisis documental posterior: la de tratar de identificar cada documento de modo que, en cualquier momento posterior, ese documento pueda ser recuperado, independientemente de que esas circunstancias contextuales no sean conocidas por el posible destinatario.

El hecho de que la información periodística pueda provenir de cualquier lugar, que esté contenida en cualquier soporte y que sea útil para múltiples y distintos usuarios, complica enormemente su manejo documental. El control del lugar en el que se generan las informaciones (que también puede ir indicado mediante distintas referencias para el mismo lugar) obliga a que también deban usarse herramientas de lenguaje controlado que formalicen las referencias geográficas incluidas dentro de los documentos.

Además, el hecho de que esa información puede estar contenida en múltiples soportes y formatos físicos complica el análisis formal de los documentos y obliga al documentalista a conocer el manejo de esos múltiples formatos y soportes para realizar el análisis de contenido.

- **Heterogeneidad:** los documentos periodísticos son muy variados en cuanto a su tamaño y en cuanto al nivel de profundidad en que tratan uno o varios temas, así como en su extensión física. Todas estas diferencias deben ser tenidas en cuenta a la hora de su procesamiento por los documentalistas. El tamaño y la profundidad son también factores que pueden determinar las peticiones de los usuarios.

Por ejemplo, si un periodista requiere datos concretos sobre un tema específico con urgencia, para complementar una información, requerirá un documento que sea breve y que contenga los detalles básicos que requiere. Si, por el contrario, ha de enfrentarse a un tema del que no conoce nada, se le requiere un reportaje en profundidad y dispone de un periodo relativamente largo, demandará todo lo que el centro de documentación pueda ofrecerle, mejor cuanto más detallado sea.

- **Redundancia:** sin duda esta característica es una de las que distinguen con mayor precisión a los documentos periodísticos. Los medios periodísticos suelen difundir, en buena parte, informaciones sobre los mismos temas en los mismos periodos de tiempo, con diversos factores añadidos que provocan fenómenos como los siguientes (BENITO AMAT y CASTILLO BLANCO, 1996, p. 189-191):
 - *Coincidencia episódica:* los medios suelen hablar de los mismos temas principales en el momento en el que se producen. Además, el mismo medio suele dar predominancia a ciertos temas, en forma de mayor atención y mayor cantidad de espacio y tiempo de difusión, durante un periodo de tiempo concreto. En el caso de las agencias, dado que su actividad productiva de noticias es continua, suelen enviar varias

versiones consecutivas de la misma noticia, añadiendo algunos detalles posteriores según los van recibiendo. Este fenómeno se repite con menor frecuencia en los medios audiovisuales y, como veremos más adelante, también en los medios accesibles a través de Internet.

- *Redundancia cíclica*: existen ciertos temas informativos que, sistemáticamente, aparecen como noticiosos en las mismas fechas a lo largo de periodos concretos. Por ejemplo, las informaciones sobre la fluidez del tráfico en las fechas de comienzo y final de los periodos de vacaciones masivos (en España: comienzo y final de los meses de verano, de las vacaciones de Navidad, Semana Santa, etc.)
- *Efecto Mateo*: aunque tomado de la comunicación científica, también sucede en el campo de la difusión periodística. Aquellos temas y autores que más han aparecido en informaciones periodísticas son los que suelen acumular las menciones de otras informaciones periodísticas. Por ejemplo, las declaraciones de los principales miembros de un Gobierno, que suelen aparecer como contenidos informativos en periodos cortos (a diario, o más de una vez a lo largo del mismo día) suelen generar múltiples referencias a los mismos, tanto directas como indirectas.
- *Modas informativas*: por diferentes razones, los medios suelen tomar un personaje o un tema de modo recurrente durante periodos concretos de tiempo. Por ejemplo, el protagonista de un hecho destacado (como un crimen) suele ser protagonista de documentos informativos difundidos durante los días sucesivos al del hecho en sí que protagonizó. Esta característica ha sido descrita por (GARCÍA GUTIÉRREZ y LUCAS FERNÁNDEZ, 1987, p. 215).
- **Dispersión**: las fuentes, los lugares, los hechos, los enfoques, los modos, los conceptos y los términos se hallan dispersos en múltiples documentos periodísticos relativos al mismo fenómeno informativo, ya sea un hecho

concreto o un tema más amplio. De hecho, Benito y Castillo apuntan la frecuencia, en la práctica periodística, de los *documentos compuestos* (que incluyen la referencia a varios hechos distintos en un mismo documento).

- **Obsolescencia:** los documentos periodísticos, debido al proceso continuo de producción periodística, pueden tener un periodo de validez informativa muy breve, aunque su valor documental permanece, como se muestra más adelante. Un fenómeno particular relacionado con esta característica es el denominado “periodo de latencia”, en el que un hecho deja de ser noticia durante un periodo indefinido, hasta que el surgimiento de nuevos detalles sobre el asunto vuelven a convertirlo en un hecho de actualidad.
- **Saturación:** el ritmo de producción de documentos periodísticos suele ser tan alto que el centro de documentación, cuya capacidad de almacenamiento es limitada, no puede en muchas ocasiones conservar todos los documentos posibles, lo cual le obliga a realizar un expurgo sistemático. Entre los criterios empleados para decidir la realización de ese expurgo se toman en cuenta, fundamentalmente, la redundancia (que permite la eliminación de documentos en favor de otros que condensan la información publicada en los eliminados) y la obsolescencia (que presupone que los documentos afectados por este fenómeno no volverán a ser útiles).
- **Ideologización:** según apunta García Gutiérrez, cada medio proyecta un código ideológico en sus periodistas, que forma parte del discurso periodístico.

2.4.2. Tipos de usuarios de los documentos periodísticos

Una vez vistas las características de los documentos periodísticos, es necesario observar cuáles son las necesidades de los posibles usuarios de esos documentos. Para ello, deben observarse en primer lugar quiénes son esos usuarios. Una tipología de los mismos puede observarse en (FUENTES I PUJOL, 1997), que distingue entre

cuatro tipos de usuarios posibles, más un quinto tipo que no contempla Fuentes pero que también forma parte de la posible tipología:

- **Periodistas:** sin duda, el tipo de usuario más habitual y frecuente de los documentos periodísticos, hasta el punto de que en muchos casos (tanto en los medios de comunicación como en parte de los estudios teóricos sobre el tema) es considerado el único.
- **Instituciones:** descritas por Fuentes como “empresas, organismos e instituciones públicas y privadas”. Fundamentalmente, requieren todo tipo de información económica y política, lo más actualizada posible, así como noticias culturales o sociales que suelen aparecer en los medios de comunicación.
- **Población académica:** entendida como el conjunto de estudiantes, profesores e investigadores de múltiples disciplinas (Fuentes cita a “historiadores, economistas, físicos, biólogos, sociólogos, médicos, politólogos, etc.”) que necesitan recoger datos de cualquier aspecto que les interese y que hayan sido recogidos en la prensa, ya sean datos muy recientes (que solo han sido recogidos por medios de comunicación), o bien, el seguimiento retrospectivo de hechos recogidos por la prensa durante un periodo concreto.
- **Público en general:** entendiendo esta categoría como cualquiera de los receptores de un medio de comunicación, que pueden pedir cualquier dato recogido por el medio, ya sea muy concreto (un artículo publicado en un periódico, por ejemplo) o lo que se ha publicado sobre un tema determinado. Este último tipo de usuarios es la que más dificultad plantea a la hora de prever sus necesidades.
- **Gestores del medio:** básicamente, nos referimos a los miembros de departamentos económicos del medio, encargados de tareas como publicidad, marketing o gestión, que deben conocer los contenidos

publicados por el medio para poder controlar su funcionamiento interno, planificar su desarrollo futuro y realizar las campañas de publicidad tanto hacia los posibles anunciantes como hacia los clientes actuales y potenciales. Como hemos citado previamente, ya Desmond en 1933 mencionaba “servir al departamento de publicidad” y “servir al departamento de difusión” como dos de las funciones de la documentación periodística.

En concreto, haremos hincapié únicamente en la primera de estas categorías de usuarios, —aunque se reconozca la necesidad de ampliar esta visión a todos los otros tipos—, dado el carácter documental-periodístico de este texto.

2.4.3. Necesidades de información de los periodistas como usuarios de documentos periodísticos

Según destacan (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 54-56), la relación entre las necesidades de información de los periodistas y la respuesta por parte del centro de documentación periodística suele ser de dos tipos:

- Por una parte, el centro de documentación se puede adelantar a las necesidades futuras de los periodistas, haciendo una previsión de temas y ofreciendo posibles documentos que solucionen esas necesidades futuras. Los motivos principales apuntados para realizar esta tarea previa son: evitar realizar estas tareas con urgencia, dado que se tiene constancia de que, tarde o temprano, habrán de realizarse; y demostrar el valor del trabajo realizado en el centro de documentación a los periodistas de la redacción, sintiéndose parte de ésta.
- Por otra parte, son los propios periodistas los que se dirigen al centro de documentación para realizar sus peticiones de información.

Básicamente, Hernández y Caridad clasifican las peticiones de los periodistas en tres categorías:

- **Peticiones indagatorias:** se trata de preguntas cortas que tratan de averiguar un dato concreto. Por ejemplo: quién es el ministro de una materia concreta.
- **Peticiones verificativas:** consisten en preguntas también cortas que pretenden comprobar datos que ya se tienen pero de cuya fiabilidad se duda. Por ejemplo, ¿la cifra que se tiene sobre las exportaciones de petróleo de un país, obtenida de otra publicación, es correcta?
- **Peticiones temáticas:** se tratan de consultas en las que se busca información de todo tipo sobre un tema. Este tipo de peticiones, las más difíciles de atender, exigen que el documentalista actúe de intermediario no ya entre el periodista y los contenidos documentales del centro de documentación, sino entre su petición inicial y lo que realmente quiere conseguir el periodista. Existen dos tipos de peticiones temáticas:
 - *Peticiones genéricas:* aquellas que se producen por desconocimiento del demandante sobre el tema del que solicita información. Por ejemplo: “todo lo que haya sobre el político XX”, que es alguien que aparece frecuentemente en los documentos periodísticos disponibles en el centro de documentación. Exige que el documentalista sepa conseguir del usuario que acote su necesidad de información (por ejemplo, información suficiente para redactar un reportaje sobre las actividades de esa persona desde que llegó al cargo).
 - *Peticiones específicas:* aquellas que se producen por parte de gente que ya tiene conocimientos del tema sobre el que solicita información, pero que quieren obtener información adicional sobre el tema. En este caso, la demanda es mucho más exigente, dado que el nivel de conocimiento previo supone altas expectativas y un juicio más riguroso sobre los

documentos que se espera recibir como respuesta. A veces, puede demandar incluso el cruce de datos obtenidos en determinados documentos. Por ejemplo: obtener los textos que contienen los reglamentos bursátiles en las bolsas europeas y de Estados Unidos, para su comparación.

Asimismo, los periodistas suelen trabajar con grandes presiones de tiempo, por lo que acostumbran a reclamar de los centros de información periodística la capacidad de ofrecer sus respuestas:

- En el menor **tiempo de respuesta** posible. Como continúan apuntando (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 43), “los medios de comunicación no admiten demoras y los periodistas exigen que sus demandas sean atendidas de forma inmediata”. Los trabajos de observación sobre prácticas concretas corroboran esta opinión:

“Journalists have to absorb information extremely quickly so invariably it has to be obtained quickly. Home and foreign news journalists are the most demanding users of all: they require facts and background data very quickly, and usually at the last minute. (...)”

It is probably sub-editors who have the least amount of time on their hands. They have a need to check, cross-check and verify information – like spellings and facts– literally within minutes.” (NICHOLAS y MARTIN, 1997, p. 48-50).

- Relacionado con lo anterior, los periodistas reclaman la recepción de respuestas en forma de documentos lo más breves y precisos posible, aplicando lo que Benito y Castillo denominan la “**ley del mínimo esfuerzo**” (BENITO AMAT y CASTILLO BLANCO, 1996). Aunque aplicado en este caso a un periódico, cualquier centro documental de prensa puede suscribir esta afirmación:

“journalists do not require comprehensiveness or exhaustivity because they always have to stop at a seemingly arbitrary point –the newspaper

deadline. So information seeking is always an impure and incomplete activity" (NICHOLAS y MARTIN, 1997, p. 47).

Es decir, los periodistas suelen trabajar habitualmente bajo la presión del reloj, con una fecha y una hora máximas para entregar su trabajo, trasladando esta urgencia a la necesidad de información demandada al centro de documentación: solo la cantidad precisa de modo que se pueda manejar en el tiempo requerido.

Para completar este apartado, aportamos, sin pretensión de agotarlas, algunas de las conclusiones concretas obtenidas por los citados (NICHOLAS y MARTIN, 1997) en el ya citado estudio de caso sobre necesidades concretas de información de periodistas demandadas al centro de documentación de un medio concreto:

- Cada sección de un medio determina una serie de peticiones de información relativamente distintas. Aquellos periodistas que se especializan en una materia concreta realizan peticiones mucho más centralizadas, mientras que quienes trabajan en áreas más amplias, como Nacional, suele ser capaces de cubrir cualquier noticia en breve tiempo, con lo que demandan datos mucho más genéricos.
- Algunas veces, los periodistas realizan peticiones generalizadas o vagas de información porque quieren evitar que otros colegas del propio medio conozcan su línea de investigación, para evitar que les "pisen" un tema, es decir, que ofrezcan una información similar al medio antes que ellos.
- Los periodistas suelen seguir un hecho antes de que se produzca, de modo que reciben documentos originales antes de que sean hechos públicos (por ejemplo, los distintos borradores de una ley). (FUENTES I PUJOL, 1997, p. 92) ahonda en esta característica desde el punto de vista del centro de documentación, recordando que los medios de comunicación son a un tiempo "productores de los textos que el documentalista ha de leer" y "usuarios de los productos que genera el documentalista".

- Los periodistas, según la opinión de Nicholas y Martin, procesan y empaquetan información, y sospechan mucho de que otro lo haga por ellos. De hecho, el citado estudio cita las sospechas, en 1997, sobre como procesa la información la base de datos de artículos periodísticos Profile, pero, por el contrario, muestra como los periodistas valoran mucho los recortes de prensa, dado que están hechos por colegas que trabajan de modo muy familiar a ellos.

2.4.4. La documentación periodística como proceso: principios

Una vez delimitado el objeto de la documentación periodística aplicable en este trabajo, los documentos periodísticos, elaborados por los medios de comunicación; establecidas las características de los mismos, y señaladas las necesidades de sus principales usuarios, los periodistas, es el turno de describir los denominados “principios operativos” (GALDÓN LÓPEZ, 2002c, p. 40-54) en los que se basa el proceso de documentación periodística para responder a los factores descritos.

Galdón distingue, inicialmente, entre dos tipos de principios:

- **Principios teleológicos:** aquellos relativos a la finalidad de la documentación periodística. Estos principios son: el verificativo, el explicativo y el editorial.
- **Principios formales y circunstanciales:** aquellos derivados del uso que se le dará a los documentos periodísticos y de las circunstancias en las que se desarrolla la actividad documental y periodística del medio en concreto. Galdón define otros tres principios de este tipo: perdurabilidad, adecuación funcional y limitativo.

Los principios teleológicos o de finalidad persiguen que la información sea verdadera, inteligible y orientadora. Pueden distinguirse así:



- **Principio verificativo:** básicamente, consiste en la comprobación de la veracidad de los textos y referencias que se valoran y se difunden, y su posterior complemento. Este principio se aplica a distintos contenidos, para cada uno de los cuales se tienen en cuenta detalles concretos:
 - *Hechos:* coincidencia con otros datos conocidos, calidad y número de fuentes citadas, distancia espacial entre el autor y el hecho, grado de elaboración y tiempo empleado, autoridad sobre la materia tratada, actualización de los datos.
 - *Juicios de valor:* grado de verdad y de integridad de los hechos sobre los que se juzga, y grado de validez de la argumentación.
 - *Opiniones:* Galdón distingue las opiniones de los juicios de valor definiendo éstas como “hipótesis acerca de las realidades sobre las que aún no hay un conocimiento cierto”. De las opiniones, se deben tener en cuenta estos aspectos: coherencia con otras opiniones anteriores del mismo autor, adecuación lógica y razonabilidad, interés ideológico o político.
 - *Ideas:* grado de novedad, entronque histórico, proyección social, efectos negativos o positivos, trascendencia...
- **Principio explicativo:** capacidad del proceso de documentación periodística para aportar explicaciones complementarias que permitan entender un hecho informativo. Este principio está lógicamente relacionado con el principio verificativo, y pretende resolver el problema creado por características de los documentos periodísticos en origen como las ya citadas heterogeneidad, redundancia o dispersión.

Al tiempo, el principio explicativo transforma la obsolescencia informativa de muchos documentos en un factor de validez que permite a estos documentos servir de base para la elaboración informativa de otros nuevos, actuando como parte de su contexto.

- **Principio editorial:** relacionado con la actividad opinativa del medio de comunicación, que no solo ofrece informaciones, sino que en muchos casos (especialmente, en medios textuales) ofrece además una serie de opiniones periódicas, no solo de editorialistas y articulistas más o menos próximos ideológicamente al medio, sino que aporta además comentarios tomados como postura oficial del propio medio, ya sea explícitamente, como en el caso de los así llamados *editoriales*, o implícitamente, a través de los citados articulistas. Ello exige que los editorialistas del propio medio sigan sus propias publicaciones, con el fin de mantener una coherencia en la expresión de las opiniones vertidas.

Por otra parte, los citados principios formales y circunstanciales, enfocados al manejo de las circunstancias que rodean al propio medio y al centro de documentación, se desglosan en los tres principios ya citados:

- El **principio de perdurabilidad**, en síntesis, sirve para definir el grado y el tiempo de permanencia y la utilidad futura de los documentos periodísticos. Sobre el establecimiento del periodo máximo de utilización de un documento periodístico como elemento de valor documental para contextualizar futuras informaciones, existen diversas opiniones al respecto.

Por ejemplo, (FUENTES I PUJOL, 1997, p. 95-96) cita un estudio realizado por Whatmore en 1978 para argumentar que los documentos con más de tres meses de antigüedad son consultados apenas por el 5% de los usuarios del centro de documentación, mientras que (NICHOLAS y MARTIN, 1997, p. 48), haciendo hincapié en la experiencia sobre los periodistas del diario británico *The Guardian*, señalaban en 1997 que un tercio de los usuarios de su servicio de documentación pedían datos con dos o más años de antigüedad.

Este principio es también entendible en el sentido de que el servicio de documentación debe ser además la memoria histórica del periódico, con el

fin de satisfacer las peticiones no solo de los periodistas, sino también de los otros tipos de usuarios citados.

- Por su parte, el **principio de adecuación funcional** consiste en la adaptación del proceso documental a cada uno de los documentos informativos, que pueden ser muy diferentes en cuanto a sus características formales (por ejemplo: los distintos tipos de trabajo periodístico y sus finalidades: noticia, reportaje en sus diferentes variedades, crónica, entrevista, artículo de opinión, editorial...) y en cuanto a sus características físicas, ya sean en cuanto al modo de difusión (texto impreso, imagen fija, gráfico, dibujo, sonido, documento audiovisual...) como en cuanto al soporte empleado (ya sea analógico, en sus distintas modalidades, o digital). Cada una de estas características determina un distinto tratamiento documental.
- Por último, el **principio limitativo** se refiere, en palabras de Galdón al “conjunto de circunstancias reales de la actividad documental periodística, que limitan su ejercicio y la utilización de medios”. En este apartado se incluye, por ejemplo, la condición del medio (no tendrá las mismas necesidades ni posibilidades un medio local que un gran medio nacional, ni los ámbitos de interés serán los mismos), las necesidades de sus distintos tipos de usuarios, la periodicidad de difusión (que hará que se valore en mayor o menor medida la capacidad de respuesta en el menor tiempo posible, por ejemplo), etc.

2.4.5. Funciones de la documentación periodística

Establecidas las características de los diferentes elementos que confluyen en la actividad documental periodística (el objeto: los documentos informativos; el sujeto: los periodistas, principalmente; y el proceso: el tratamiento de los documentos periodísticos producidos por el medio), pueden delimitarse sus funciones. En este

punto, aunque existen otras menciones ya señaladas por diversos autores, seguiremos por su mayor detalle la definición de (GALDÓN LÓPEZ, 2002, p. 67-88), que recoge y amplía otras similares dadas en (GALDÓN LÓPEZ, 1986, p. 24-27), y (GALDÓN LÓPEZ, 1994, p. 170-195).

Estas funciones pueden agruparse en varias categorías: las funciones destinadas al conocimiento de la realidad (distinguidas por Galdón en 1986): *previsoria*, *preparatoria*, *crítico-verificadora* y *completiva*), las dirigidas a mejorar la comunicación en la elaboración de textos periodísticos (*lingüístico-clarificadora*, *modélico-narrativa* e *informativa propia*) y otras dos dirigidas a mejorar actividades profesionales de los periodistas (*orientadora* y *profesional*), así como lo que Galdón denomina una “propuesta” (la función *rectificadora*).

La primera de estas categorías, formada por las funciones para el conocimiento de la realidad, está compuesta por:

- La **función previsoria** (*sic*), que consiste en ofrecer a los periodistas información sobre hechos que pueden suceder, bien periódicamente (recordemos el fenómeno de la redundancia cíclica), bien porque están previstos en un periodo concreto (por ejemplo, un acontecimiento deportivo), bien como desarrollo de otros hechos que se están sucediendo (por ejemplo, en el caso de una enfermedad terminal de un personaje considerado como noticiable, una biografía que se publicará en el momento de su fallecimiento).
- La **función preparatoria** consiste, en palabras de Galdón, en
“proporcionar a cada informador la documentación que necesita en cada caso concreto antes de enfrentarse con el objeto informativo que debe abordar. (...) Esta preparación documental previa es el elemento más importante en la realización de cualquier semblanza, entrevista o reportaje” (GALDÓN LÓPEZ, 2002, p. 70).
- La **función crítico-verificadora**, desarrollada por Galdón en tres dimensiones: la *fáctica*, o comprobación de la exactitud de los hechos,

nombres, direcciones, citas, etc.; la *crítica*, que pretende indagar si las afirmaciones vertidas por todo tipo de personas o fuentes son verdaderas o falsas; y la *semántica*, que pretende mostrar la correspondencia entre lo que se pretende comunicar y el lenguaje empleado para su divulgación.

- La **función completiva** (también conocida como **explicativa**), que sirve para contextualizar los acontecimientos, dándoles una perspectiva en el tiempo (en un periodo histórico concreto, breve o amplio) y en el espacio (en su ubicación geográfica), estableciendo las causas de un hecho o un tema informativos. Esta función se aplica a todo tipo de fuentes. Galdón señala como caso especial las fuentes iconográficas, que permiten a veces una explicación sintética pero muy eficaz.

A continuación, Galdón señala tres funciones referidas a mejorar la elaboración de informaciones con vistas a su comunicación a los receptores:

- La **función lingüístico-clarificadora**, que trata de permitir a los periodistas elaborar sus informaciones de modo que resulten más claras y precisas para el público en general, frente a la complejidad forzada y a la oscuridad del lenguaje que se recibe en toda una serie de documentos externos y de declaraciones públicas por parte de técnicos y también de políticos. Esta función se desarrolla sobre todo mediante el adecuado uso de todo tipo de fuentes lingüísticas (sobre todo, diccionarios).
- La **función modélico-narrativa** consiste en

“disponer de formas de expresión claras y variadas en cuanto a estructura y estilo que, en un momento concreto, puedan facilitar el trabajo de los informadores cuando deban comunicar ideas o realidades complejas y difícilmente comunicables porque precisen un lenguaje o una estructura narrativa especial” (GALDÓN LÓPEZ, 2002, p. 82).

Esta opción serviría para permitir a los informadores darles ideas que permitan mejorar el modo de difundir sus informaciones más complejas, aunque, como apunta el propio Galdón, ni los propios medios ni los

teóricos de la documentación periodística han hecho aún mucho hincapié en el tema.

- La **función informativa propia** consiste en la elaboración, por el propio servicio de documentación, de piezas informativas que sean incluidas tal cual en el medio, a diferencia de la práctica en la que el centro documental aporta datos que luego reelaboran los periodistas. Esta práctica se ha desarrollado en los medios en forma de obituarios (o semblanzas de personajes fallecidos que, en un momento determinado de la historia reciente, fueron importantes), semblanzas biográficas, recuadros verificativos de declaraciones personales o explicativos de temas complejos, cronologías, recuadros de antecedentes, agendas, dossiers o informes a fondo sobre temas concretos, recuadros referenciales e infografías.

En realidad, en nuestra opinión, esta función no es más que la aplicación de las otras funciones (en particular, de la explicativa), de un modo directo, por parte del servicio de documentación. Su importancia radica en la asunción de la importancia que el centro documental tiene, hasta el punto de ofrecerle la posibilidad de que difunda la importancia de su existencia no solo a la propia redacción, sino directamente a los receptores finales. De hecho, ya en 1984, Albert citado por (GALDÓN LÓPEZ, 2002, p. 83) señalaba que “la informatización de los servicios de documentación periodística permitirá, con pocos gastos y sin un esfuerzo particular, multiplicar esos dossiers documentales anejos y extender el servicio a usuarios externos”. Este elemento será desarrollado con mayor profundidad en la presente tesis.

Finalmente, se mencionan otras funciones que pueden agruparse como tendentes a la mejora de la actividad profesional periodística:

- La **función profesional**, consistente en la recopilación y aportación por el servicio de documentación periodística de todo tipo de obras que sirvan a

los periodistas del medio correspondiente para mejorar sus prácticas profesionales de todo tipo, siguiendo la pauta del periodista polaco Kapuscinski, citado por (GALDÓN LÓPEZ, 2002, p. 87), de que “es parte inherente a la profesión periodística... la actualización y el estudio constantes”.

- La **función rectificadora**, propuesta por Galdón como una forma de que el propio servicio de documentación revise los propios contenidos del medio, localizando los posibles errores cometidos sistemáticamente, e institucionalizando incluso una sección en el medio que los muestre públicamente, más allá de la ocasional “fe de erratas”.

2.4.6. Actividades documentales sobre la información periodística

¿Cómo aplicar las funciones de la documentación periodística sobre el conjunto de documentos periodísticos, con el fin de que sean útiles a sus usuarios? El elemento mínimo sobre el que actúa un centro de documentación periodística es la llamada **unidad documental**, que Cid, parafraseando la definición genérica al respecto de (GUINCHAT y MENOUE, 1992, p. 96), define como la

“fuente de información, sea ésta un documento completo o una parte del documento aislada artificialmente (p. ej., un artículo de un periódico), considerada como una unidad informativa en sí misma, y que como tal es la base de una sola descripción documental” (CID LEAL, 1995, p. 91).

En el caso de la prensa, al menos en su vertiente tradicional, se ha considerado habitualmente que “el diario (...) es formato de multitud de artículos e informaciones que constituyen, cada una, una unidad documental” (COTTE, 1991, p. 13, citada en FUENTES y CONESA, 1994, p. 21).

Sobre estas unidades documentales se aplica lo que la citada (CID LEAL, 1995, p. 92) denomina la **cadena documental**. Cid distingue tres fases principales en la cadena documental, cada una de las cuales incluye la realización de una serie de

actividades: la fase de **entrada** de los documentos periodísticos en el centro documental, que comprende las operaciones de *selección, adquisición y registro*; la fase de **tratamiento y análisis** de esos documentos, que comprende el *análisis documental* (tanto *descriptivo* como *de contenido*) y el *procesamiento técnico*; y finalmente, la fase de **salida** de los documentos informativos, ya procesados, como respuesta a las peticiones de los usuarios, que comprende las actividades de *recuperación y difusión*.

En este trabajo, nos referiremos a estas tareas aplicadas a los documentos producidos por el propio medio, que han de ser registrados, analizados y recuperados por el centro de documentación del mismo.

2.4.6.1. La entrada de los documentos

La fase de entrada de los documentos periodísticos en el centro documental comprende, a su vez, varias fases diferentes:

- La **selección** de aquellas fuentes en las que se realizará la búsqueda de los documentos que el centro de documentación considere adecuados para cumplir las citadas funciones de la documentación periodística¹, teniendo en cuenta que deben cumplirse dos premisas básicas:
 - Localizar los documentos que satisfagan las necesidades de los usuarios futuros.

¹ Conviene, en este punto, recordar que el concepto de “selección”, según apunta (LÓPEZ HERNÁNDEZ, 1999), no se produce solo en esta fase de entrada, sino en todas las fases del proceso documental. Así, para López Hernández, la indización (en la fase de análisis) es una selección de contenido temático, mientras la recuperación de la información supone a su vez dos procesos de selección: por una parte, de las fuentes en las que se puede buscar (por ejemplo, las bases de datos); por otra, la selección de aquellos documentos que interesen realmente al usuario final. En todo caso, en el marco de la presente tesis, consideraremos con el término de “selección” solo a la primera acepción, la aplicada en la fase de entrada.

- Considerar el peligro de saturación del centro documental, debido especialmente a dos características negativas de los documentos periodísticos ya citadas: la redundancia y la obsolescencia.

(LÓPEZ HERNÁNDEZ, 1999, p. 231-233) señala como criterios de selección por parte de contenidos, por parte de un medio periodístico, los siguientes: el valor económico de la fuente sobre la que se realiza la selección (en el caso de los documentos elaborados por el medio, equivale a la decisión sobre si materialmente se puede conservar o no, y qué beneficios obtendrá el medio de su conservación), la importancia informativa básica de los mismos, y finalmente, su grado de importancia temporal (a medio y largo plazo).

- La **adquisición**, que consiste en la de obtención física de los documentos seleccionados.
- El **registro**, consistente en la identificación inequívoca de cada documento periodístico, con una doble función, según (CID LEAL, 1995, p. 96): el control de las pérdidas de documentos y la realización de un inventario económico que sirva al departamento de gestión del medio de comunicación para controlar gastos económicos.

2.4.6.2. El análisis formal

La segunda fase documental, el **análisis documental** consiste en el procesamiento de los documentos periodísticos, físicamente ya identificados, con el fin de generar herramientas documentales que permitan su conservación futura y su recuperación.

Esta fase se puede dividir a su vez en otra dos: el **análisis formal**, que consiste en la definición de las características formales del documento, especialmente útil cuando se dispone de varios formatos distintos para los documentos recibidos, con el fin de facilitar su reproducción; y el **análisis de contenido**, que consiste en la

representación intelectual de los contenidos de los documentos, con el fin de facilitar su recuperación.

En el análisis formal se incluye la identificación de los datos de publicación o de emisión de cada una de las unidades documentales producidas por el centro. Centrándonos en el ámbito textual, podrían extraerse de modo genérico, entre otras: el título del documento, el tipo de documento (y en particular, si fuera posible, su género), la sección y el área concreta de la misma en la que se incluyó la unidad documental al publicarse, la mención de responsabilidad (el autor o autores, los editores, los colaboradores, etc.), la designación del medio en el que se ha hecho público (si el centro de documentación formara parte de un mismo medio que realice varias publicaciones, o si forma parte de un grupo en el que se almacenaran los contenidos de distintos medios), la descripción física, si forma parte de un soporte específico, y otras notas varias.

2.4.6.3. El análisis de contenido

Más complejo resulta en un centro documental periodístico el **análisis de contenido**, entendido como el conjunto de procedimientos empleados que

“consiste en captar o percibir los significados esenciales del contenido documental a fin de representarlos adecuadamente para que, entre otras cosas, sirvan como elementos de recuperación” (GARCÍA GUTIÉRREZ, 1999, p. 285).

Esos “significados esenciales” son, para (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 51)

“aquellas partes, personas físicas y jurídicas, objetos, temas, hechos, situaciones, acciones, lugares y fechas por los que más tarde un periodista pueda desear recuperar información. El fin último del proceso de lectura y análisis de contenido es la síntesis del texto”.

La representación documental de la información periodística debe enfrentarse con las características de los documentos periodísticos, que hemos reseñado anteriormente;

pero también, con las necesidades de los principales usuarios de esos documentos: los mismos periodistas.

Derivadas de ambos factores, a la hora de representar las unidades documentales en un centro de documentación periodístico, surgen dificultades como las siguientes:

- a) El centro de documentación, por definición, debe enfrentarse a una velocidad de respuesta lo más alta posible, dado que sus principales usuarios son periodistas que, como acabamos de señalar, trabajan bajo una presión temporal continua.
- b) Características como el enciclopedismo, la universalidad y la heterogeneidad dificultan el empleo de mecanismos de representación documental para organizar todos esos documentos en conjunto, dado la gran variedad de temas y de referencias (temáticas, temporales, personales, etc.) por las que se puede solicitar un documento.
- c) Características como la redundancia o la dispersión obligan a tratar de relacionar los contenidos del centro de documentación, de modo que los acontecimientos puedan ser entendidos: un mismo acontecimiento suele estar referido en varios documentos, y muchos acontecimientos pueden tener una relación que ayude a su explicación. Más aún, debe contarse con que la obsolescencia de los acontecimientos citados puede suponer que hechos o términos dados por supuestos a la hora de elaborar el texto no sean comprendidos en el futuro sin una explicación previa.
- d) Por otra parte, los documentos periodísticos textuales, en el caso que nos ocupa, presentan también, en el lenguaje natural empleado para su elaboración, los problemas propios del análisis de contenido en general, que pueden encontrarse descritos, por ejemplo, en (LANCASTER, 1995): la sinonimia y la cuasisinonimia, la ambigüedad de ciertos términos sin un calificador anexo, la homografía y la homonimia, entre otros.

A ellos, se unen, sin pretensión de agotar los posibles problemas lingüísticos derivados del lenguaje libre, otros fenómenos como las paráfrasis o las ambigüedades intencionadas, como señalan (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 52), así como las omisiones en el texto, cuyos motivos y funciones han sido tipificados por (GARCÍA GUTIÉRREZ, 1992), sin olvidar la influencia que, sobre la comprensión de un texto, puede tener el contexto que lo rodea en sus distintas variedades, descritas entre otros por (VAN DIJK, 1990).

Para solucionar estos problemas, se plantea una disyuntiva entre el uso de lenguajes libres y la aplicación de lenguajes controlados para describir los documentos periodísticos; esta doble posibilidad ha sido tratada en diversos trabajos. Entre ellos, (RODRÍGUEZ VELA, 1992) señala las ventajas e inconvenientes de cada uno, indicando que los lenguajes libres poseen una mayor riqueza terminológica y ayudan a expandir el contenido de las búsquedas, además de no suponer un coste adicional en su generación, aunque sí el mantenimiento de un fichero invertido muy elevado.

En cuanto a los lenguajes controlados, se consideran especialmente útiles en documentación periodística, dado que se trata de un campo de terminología inestable y poco normalizada (GARCÍA GUTIÉRREZ y LUCAS FERNÁNDEZ, 1987, p. 68). El uso de lenguajes controlados, además de reducir y hacer más estable la terminología empleada, permiten determinar diferentes tipos de relaciones entre los términos empleados en la descripción, con el fin de mejorar la recuperación de los documentos indicando a los usuarios futuros qué terminos emplear, y cuáles no pueden emplearse.

No obstante, aplicando el principio de adecuación funcional (en palabras del antes citado Galdón) al medio en el que se integra, la descripción realizada en el centro de documentación debe contar con otros condicionantes: los factores económicos (HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 52) y asimismo, la continua multiplicación de información heterogénea (FUENTES I PUJOL, 1997) que absorbe continuamente el centro.

Repasando los instrumentos documentales empleados habitualmente por los centros de documentación, podemos encontrarnos con los siguientes:

- **Archivos de recortes:** uno de los instrumentos de descripción documental más antiguos. Como ya se ha apuntado, consiste en la elaboración de carpetas, cada una de ellas encabezada mediante un término incluido en una clasificación, habitualmente diseñada por el propio medio. Dado que un mismo documento puede referir a más de un tema, se incluirán en el archivo tantas copias del documento respectivo como posibles temas incluidos en el documento tengan su correspondiente carpeta.

Esta fue una de las primeras herramientas documentales en medios de comunicación, como ya hemos señalado en la breve sinopsis histórica sobre la documentación periodística. Algunas de las ventajas que ofrecen los archivos de recortes son las siguientes (FUENTES y CONESA, 1994):

- Están escritos de una forma inteligible para el periodista, con el mismo estilo y el mismo formato que usa para elaborar sus trabajos.
- Están ordenados de modo que lo último en publicarse sobre un tema es lo primero que aparece
- Son fácilmente transportables y copiables.

Sin embargo, también presentan dos graves problemas: en primer lugar, contribuyen notablemente a la saturación física del centro de documentación. Y además, “la baja calidad del papel prensa acelera su deterioro” (FUENTES I PUJOL, GONZÁLEZ QUESADA, y JIMÉNEZ LÓPEZ, 2000, p. 394).

Además, al ser herramientas de clasificación, su carácter rígido no permite enfrentarse correctamente a la redundancia de la producción periodística. Es más, al tener que enfrentarse con el fenómeno de la dispersión, entendida también como la existencia de varios temas en un mismo

documento periodístico, provocan aún mayor redundancia en el centro documental, al requerirse físicamente varias copias de una misma noticia para almacenarla en tantas carpetas como términos de descripción básicos se hayan utilizado para la misma.

- **Índices:** constituyen una herramienta de localización de los documentos en el centro de documentación. Para su elaboración, se seleccionan los términos que pueden emplearse para encabezar las carpetas en las que se introducen los documentos, incluidos los nombres de los principales personajes que aparecen dentro de los documentos. En todo caso, el número de términos extraídos no suele ser elevado. En palabras de (SEMONCHE, 1993),

“the number of terms used is dependent upon the contents of the article: three of four relatively specific terms per article are not uncommon, while some local articles require even more. These terms are in addition to the personal names assigned to the articles”.

Estos índices se han ido automatizando durante la década de los 90 al menos en aquellos centros documentales de medios con cierto poder adquisitivo, como recogen, al menos en nuestro país, diversos trabajos; por citar algunos, (SALMURRI, 1994), (VANACLOCHA, 1994), (MAZORRA, 1994), (MÁRQUEZ, 1994), (SERRANO y COLLA, 1994) o (CANTERA, 1995).

En otros casos, y especialmente en aquellos índices concebidos como una herramienta de acceso público (por ejemplo, (EL PAÍS, 1989), publicado anualmente desde 1977), el índice es agrupado empleando los titulares, ordenándolos por temas o por fechas de publicación.

- **Fichas:** concebidas como descripciones de catalogación y de indización a un tiempo: puede verse un ejemplo de su descripción en (SERRANO y COLLA, 1994). Contienen las referencias formales básicas, ya descritas, así como el empleo de los descriptores asociados a cada documento, realizados

mediante el *vaciado*, o extracción de todos los documentos del ejemplar original. En muchos casos, también a partir de los 90, se realizaron diversas bases de datos referenciales que contenían estas fichas, con el fin de reducir el tamaño requerido y facilitar su recuperación mediante programas de recuperación que utilizaban operadores booleanos, principalmente.

- **Ficheros de autoridades:** término amplio que agrupa, según la citada (SEMONCHE, 1993), otros más específicos como lista de autoridades, lista de palabras clave, etc. y que en otros medios se ha dado también en llamar “diccionario” (*sic*) (MARTÍN, 1994). En algunos casos, se ha incluido también, erróneamente, a los tesauros en esta clasificación.

La profundidad de este tipo de documentos, varía mucho de unos medios a otros, como puede observarse de las referencias antes citadas sobre la prensa española en los 90 y de otros casos que pueden encontrarse, por ejemplo, en las ponencias recogidas por (BELLVESER, 1994), o en (PASTOR RUIZ, 1992).

- **Tesauros:** se trata de lenguajes documentales que permiten establecer relaciones de equivalencia, jerarquía y asociación entre los términos utilizados para la descripción documental. Existen pocos casos de uso de tesauros entre los medios de comunicación, dado que su mantenimiento resulta difícil de compaginar con las citadas características de la documentación periodística, y en particular, con la universalidad. No obstante, algunos medios sí realizan un esfuerzo por mantener los ya existentes o por elaborar otros nuevos.
- **Resúmenes:** fundamentalmente, son de tipo indicativo, esto es, tienen como función la “breve y exacta representación del contenido que como primer objetivo y sobre cualquier otro tópico debe reflejar la macroestructura general” (MOREIRO, 1993, p. 149). En este caso, debe entenderse la macroestructura general como “una proposición formada por la secuencia

de proposiciones contenida en un discurso o en parte de él” (MOREIRO, 1993, p. 54).

Los resúmenes se realizan especialmente cuando se pretende ofrecer los contenidos accesibles a público externo, o bien, como parte de las citadas fichas.

- Las **bases de datos**, que pueden ser referenciales, o bien, incluir todo el contenido publicado por un medio a texto completo, con el fin de facilitar su recuperación mediante diversas opciones, como el uso de los citados índices, ficheros de autoridades, operadores, etc. En este caso, pueden incluir o no, además del texto, algunos campos de descripción formal y de análisis documental.

2.4.6.4. La organización y conservación de los fondos

Realizado el análisis documental, los documentos y sus representaciones documentales (las herramientas arriba citadas) deben someterse a actividades de **ordenación, almacenamiento y conservación**.

En cuanto a la ordenación, como en cualquier otro centro documental, cabe ofrecer por válido el siguiente criterio (GUINCHAT y MENOUE, 1992, p. 96):

“La ordenación que se elija debe permitir localizar dichas unidades de forma rápida y fácil, y para ello debe ser lo más sencilla e inteligible posible para los usuarios; cada unidad deberá tener un lugar único, siempre el mismo; poseer facilidades de ampliación, debe permitir detectar errores de ubicación y, por último, garantizar una buena conservación de los documentos”.

Las principales formas de almacenar y conservar las unidades documentales periodísticas y, en su caso, las descripciones realizadas, pueden dividirse en:

- Los citados **archivos de recortes**, también denominados por (VALLE GASTAMINZA, 1995, p. 72) **dossiers**, que pueden estar agrupados

nominalmente por áreas geográficas (países, regiones, etc.), por temas y personajes.

- Las versiones de los **índices** que son consultables, a través del acceso mediante ordenadores en línea, a través de soportes ópticos, como el CD-ROM, o bien, a través de su versión en papel (aunque esta última opción está en desuso).
- Las mencionadas **bases de datos**, almacenadas en diversos soportes.
- Las **colecciones de periódicos**, que permiten el acceso cronológico y sirven para que el centro de documentación sirva como memoria histórica del propio medio.

Estas colecciones pueden estar almacenadas en diversos soportes: ya sea en papel, o en los mencionados soportes micrográficos, como el **microfilm** o la **microficha**; o bien, en soportes ópticos; o finalmente, almacenados en soportes magnéticos que, por ejemplo, permitan su consulta en línea. En estos dos últimos casos, los contenidos consultados pueden generarse mediante el escaneado previo del original y conversión al formato digital (por ejemplo, PDF), o, en el caso de los nuevos medios, mediante copia del original en diversos formatos.

Un caso concreto de esta posibilidad son los citados **sistemas de gestión electrónica de documentos (GED)**: según (FUENTES I PUJOL, GONZÁLEZ QUESADA, y JIMÉNEZ LÓPEZ, 2000, p. 393-394),

“en los sistemas GED las imágenes digitalizadas de los artículos, almacenadas en discos ópticos, se relacionan con una base de datos referencial que indiza cada documento. Este sistema sustituye la microfilmación del dossier por un formato accesible informáticamente. Simplifica la captura del documento, que se realiza mediante un escáner, así como la consulta. Aporta también soluciones a los principales problemas de gestión de los dossiers de prensa en los servicios de documentación: ahorro de espacio físico, preservación del material... y rapidez de acceso al documento”.

2.4.6.5. La difusión documental: hacia una nueva visión a partir de los contenidos digitales

Finalmente, el centro de documentación, de acuerdo a las necesidades expresadas por sus usuarios (de los cuales el grupo más habitual, el de los periodistas, ha sido explicado en un punto anterior del presente capítulo) emplea tanto las referencias y herramientas documentales como los propios documentos, para difundir la información recopilada, cumpliendo con sus citadas funciones documentales.

Hasta no hace mucho, e incluso aún en buena parte de los medios periodísticos, el centro de documentación es quien realiza la función de búsqueda de datos concretos para todo tipo de necesidades específicas. A ello se une la elaboración de otros productos documentales, como las agendas de acontecimientos, que sirven para establecer los hechos previstos en un periodo concreto. Estas agendas sirven al centro de documentación para establecer un calendario de trabajo que ayude a una mejor planificación del mismo, al determinar tareas a realizar obligatoriamente (la recopilación de información sobre esos acontecimientos) en un periodo concreto, previo o simultáneo al desarrollo de estos hechos.

No obstante, el manejo cada vez mayor de herramientas informáticas ha supuesto que, al menos en los grandes medios, los usuarios mayoritarios, los periodistas, hayan comenzado a acceder directamente a estos contenidos almacenados en soportes digitales, así como a los diversos productos elaborados por el centro de documentación en su función informativa propia (como las mencionadas agendas).

Este acceso, sin embargo, suele ser más sencillo cuando la información se refiere a datos concretos, fácilmente localizables. En el caso de que la información exija una mayor complejidad, se requiere una consulta directa al centro de documentación, aunque en muchos casos, esa consulta se pueda realizar a través de herramientas telemáticas, como el correo electrónico, y no sea precisa la consulta directa al centro, en muchos casos.

El capítulo siguiente muestra como la representación documental periodística mediante herramientas digitales y la posibilidad del acceso en línea al centro de

documentación fomenta especialmente el empleo de nuevas formas de difusión documental. El desarrollo de estas nuevas formas puede afectar, como desarrollaremos en capítulos posteriores, al modo de trabajo de los centros de documentación periodística, e incluso, al propio concepto del centro.

Referencias

AFP. *AFP Historial. Un poco de historia*. 2003. [Consulta: 10/8/2003].
<<http://www.afp.com/espanol/afp/?cat=history>>

AGUADO, Javier. El centro de documentación y archivo de Prensa Española. En: *Boletín de la SEDIC*, 1995, nº 21, p. 2.

ALBERT, Pierre. *Historia de la prensa*. Madrid: Rialp, 1990.

ARMENTIA, José Ignacio, CAMINOS, José María, ELEXGARAY, Jon, MARTÍN, Flora, y MERCHÁN, Iker. *El diario digital. Análisis de los contenidos textuales, aspectos formales y publicitarios*. Barcelona: Bosch, 2000.

ASSOCIATED PRESS. *The AP's origins*. 2003. [Consulta: 10/8/2003].
<<http://www.ap.org/pages/history/origins.htm>>

BELLVESER, Enric. II Seminari La Documentació als mitjans d'informació :l'experiència multimèdia : ponències i conclusions. En: *II Seminari La Documentació als mitjans d'informació :l'experiència multimèdia : ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana. Direcció General de Mitjans de Comunicació Social, 1994.

BENITO AMAT, Carlos y CASTILLO BLANCO, Lourdes. Factores condicionantes de las operaciones documentales en medios de comunicación. En: *Revista Española de Documentación Científica*, 1996, vol. 19, nº 2, p. 188-201.

CANTERA, Gregorio. Centro de documentación del grupo Hachette Filipacchi. En: *Boletín de la SEDIC*, 1995, nº 21, p. 7.

CARIDAD, Mercedes. *La teledocumentación*. Madrid: Forja, 1984.

CID LEAL, Pilar. La cadena documental y su aplicación a la documentación periodística. En: FUENTES i PUJOL, M^a Eulàlia (editora). *Manual de documentación periodística*. Madrid: Síntesis, 1995, p. 91-109.

CODINA, Lluís. La Documentación en los medios de comunicación: situación actual y perspectivas de futuro. En: *I Congreso Universitario de Ciencias de la Documentación. Teoría, historia y metodología de la Documentación en España (1975-2000)*, Universidad Complutense de Madrid, Facultad de CC. de la Información. Madrid: Dpto. de Biblioteconomía y Documentación, Universidad Complutense de Madrid, 2000, [Consulta: 6/3/2003].
<<http://www.ucm.es/info/multidoc/multidoc/revista/num10/paginas/pdfs/Codina.pdf>>

COLL-VINENT, Roberto. La documentación de prensa en España. En: *Documentación de las Ciencias de la Información*, 1978, vol. 2, p. 25-31.

COTTE, Dominique. *Stratégie documentaire dans la presse*. París: ESF, 1991.

DELGADO SOLER, Cinta. Documentación de prensa. En: GARCÍA GUTIÉRREZ, Antonio (editor.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999, p. 449-472.

El País. Índice del año 1988. Madrid: PRISA, 1989.

FOLEY, Katty y BRISCOE, Ellen O. The newspaper library in the information age: a personal view from within. En: *Online*, 1989, vol. 13, n^o 6, p. 15-24.

FUENTES I PUJOL, Eulàlia, GONZÁLEZ QUESADA, Alfons, y JIMÉNEZ LÓPEZ, Àngels. Documentación e información electrónica. En: MOREIRO, José Antonio (Coordinador.). *Manual de documentación informativa*. Madrid: Cátedra, 2000, p. 345-421.

FUENTES i PUJOL, M^a Eulàlia. *Servicio documental en la prensa diaria: análisis y orientaciones*. Barcelona: Mitre, 1984.

FUENTES I PUJOL, M. Eulàlia. Evolución de la documentación periodística a España durante los últimos cinco años y algunas experiencias europeas. En: *II Seminario La Documentación en los medios de información: la experiencia multimedia: ponencias y conclusiones*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 17-28.

FUENTES I PUJOL, M^a Eulàlia y GONZÁLEZ QUESADA, Alfonso. La prensa española en Internet. Análisis de los servicios de valor añadido. En: *Los sistemas de información al servicio de la sociedad. Actas de las VI Jornadas Españolas de Documentación*, 29 al 31 de octubre de 1998. Valencia: Fesabid, 1998, p. 281-292.

FUENTES i PUJOL, María Eulàlia. *Documentación y periodismo*. Barañáin: EUNSA, 1997.

FUENTES, M. Eulàlia y CONESA, Alicia. *La documentación periodística. Catalunya, España y otras experiencias europeas*. Barcelona: Generalitat de Catalunya, Centre de Investigació de la Comunicació, 1994.

GALDÓN LÓPEZ, Gabriel. *El servicio de documentación de prensa: funciones y métodos*. Madrid: Dossat, 1986.

----. *Principios operativos de la documentación periodística*. Madrid: Dossat, 1989.

----. *Desinformación. Método, aspectos y soluciones*. Barcelona: Ariel, 1994.

----. Funciones informativas de la documentación periodística. En: GALDÓN, Gabriel (coord.). *Teoría y práctica de la documentación informativa*. Barcelona: Ariel, 2002a, p. 67-88.

----. *Perfil histórico de la documentación en la prensa de información general (1845-1984)*. 4ª. Pamplona: Ediciones de la Universidad de Navarra, 2002b.

----. Principios y criterios operativos. En: GALDÓN, Gabriel (coord.). *Teoría y práctica de la documentación informativa*. Barcelona: Ariel, 2002c, p. 31-65.

GARCÍA GONZÁLEZ, Gloria. La conformación de la moderna prensa informativa (1848-1914). En: GÓMEZ MOMPART, Josep Lluís y MARÍN OTTO, Enric (editores.). *Historia del periodismo universal*. Madrid: Síntesis, 1999, p. 49-99.

GARCÍA GUTIÉRREZ, Antonio. *Análisis documental del discurso periodístico*. Madrid: Centro de Técnicas Documentales, 1992.

----. Análisis documental: textos periodísticos escritos. En: GARCÍA GUTIÉRREZ, Antonio (editor.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999a, p. 285-303.

----. Proceso documental y discurso periodístico: modelo teórico y procedimientos generales. En: GARCÍA GUTIÉRREZ, Antonio (editor.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999b, p. 211-222.

GARCÍA GUTIÉRREZ, Antonio y LUCAS FERNÁNDEZ, Ricardo. *Documentación automatizada en los medios informativos*. Madrid: Paraninfo, 1987.

GONZÁLEZ QUESADA, Alfons. La evolución histórica de la documentación periodista. En: FUENTES i PUJOL, Mª Eulàlia (editora.). *Manual de documentación periodística*. Madrid: Síntesis, 1995, p. 23-39.

GUINCHAT, Claire y MENOUE, Michel. *Introducción general a las ciencias y técnicas de la información y documentación*. Madrid: CINDOC, UNESCO, 1992.

HERNÁNDEZ PÉREZ, Antonio y CARIDAD SEBASTIÁN, Mercedes. Documentación escrita en los centros de documentación de los medios de comunicación. En: MOREIRO, José Antonio (Coordinador.). *Manual de documentación informativa*. Madrid: Cátedra, 2000, p. 37-82.

JIMÉNEZ LÓPEZ, Àngels. *Estudi de la gestió documental de la informació en els serveis de valor afegit dels mitjans de comunicació a internet: el cas de la premsa diària a l'estat espanyol*. Universidad Autònoma de Barcelona, 2001. [Consulta: 2/12/2003]. <<http://www.tdcat.cbuc.es/TDX-1008102-143437/index.html>>

LANCASTER, Frederick Wilfrid. *El control del vocabulario en la recuperación de información*. Valencia: Universitat de València, 1995.

LÓPEZ DE ZUAZO ALGAR, Antonio. *Diccionario del periodismo*. Madrid: Pirámide, 1977.

LÓPEZ HERNÁNDEZ, M^a Ángeles. La selección de documentos. En: GARCÍA GUTIÉRREZ, Antonio (editor.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999, p. 223-241.

LÓPEZ YEPES, José y ROS GARCÍA, Juan. *¿Qué es Documentación? Teoría e historia del concepto en España*. Madrid: Síntesis, 1993.

MÁRQUEZ, Manuel. El tratamiento de la información deportiva. En: *II Seminari La Documentació als mitjans d'informació: l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 73-76.

MARTÍN, Mauricio. El Correo. Documentación gráfica. Hasta el archivo digital. En: *II Seminari La Documentació als mitjans d'informació: l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 53-63.

MARTIN, Shannon E. y HANSEN, Kathleen A. *Newspapers of record in a digital age. From hot type to hot link*. Westport: Praeger, 1998.

MAZORRA, Elena. Principios y objetivos de la informatización del Servicio de Documentación de la revista Tiempo. En: *II Seminari La Documentació als mitjans d'informació :l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 87-92.

MOREIRO, José Antonio. *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III de Madrid - Boletín Oficial del Estado, 1993.

NICHOLAS, David y MARTIN, Helen. Assessing information needs: a case study of journalists. En: *Aslib Proceedings*, 1997, vol. 49, n° 2, p. 43-52.

PASTOR RUIZ, Fátima. Problemática y tendencias sobre la información de los archivos periodísticos. En: *Revista General de Información y Documentación*, 1992, vol. 2, n° 1, p. 99-108.

PROVENZANO, Dominic. Nexis. En: *Database*, 1981, vol. 4, n° 4, p. 30-41.

RODRÍGUEZ VELA, Cristina. Los lenguajes documentales en las bases de información política y de actualidad. En: *Revista Española de Documentación Científica*, 1992, vol. 15, n° 1, p. 13-23.

SALMURRI, Carles. L'automatització del Servei de Documentació i Arxiu de "La Vanguardia". En: *II Seminari La Documentació als mitjans d'informació :l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 35-40.

SEMONCHE, Barbara P. *Newspaper indexing policies and procedures*. 1993. [Consulta: 25/6/2002].
<<http://www.ibiblio.org/journalism/indexing.html>>

-----, *History of News Libraries*. 2003. [Consulta: 10/9/2003].
<<http://parklibrary.jomc.unc.edu/newslibhist2.html>>

SERRANO, Rosa y COLLA, Rosari. Criteris d'anàlisi en el banc de dades Datemps. En: *II Seminari La Documentació als mitjans d'informació: l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 77-86.

TAPIA LÓPEZ, Alicia. Antecedentes de la documentación periodística en España. En: *Documentación de las Ciencias de la Información*, 1995, nº 18, p. 233-244.

VALLE GASTAMINZA, Félix del. La documentación de los medios informativos. En: EL MIR, Amado José y VALBUENA DE LA FUENTE, Felicísimo (compiladores.). *Manual de Periodismo*. Las Palmas; Barcelona: Universidad de las Palmas de Gran Canaria; Prensa Ibérica, 1995, p. 61-83.

VAN DIJK, Teun A. *La noticia como discurso. Comprensión, estructura y producción de la información*. Barcelona: Paidós, 1990.

VANACLOCHA, José. El servicio de documentación escrita en Cartelera Turia. En: *II Seminari La Documentació als mitjans d'informació: l'experiència multimèdia: ponències i conclusions*, Valencia. Valencia: Generalitat Valenciana, 1994, p. 65-72.

Capítulo 3

EL PERIODISMO DIGITAL EN LÍNEA Y LA DOCUMENTACIÓN PERIODÍSTICA

3.1. Introducción

El periodismo en línea, entendiéndolo por tal en el contexto del presente trabajo la difusión de documentos periodísticos digitales a través de Internet y, en particular, mediante la World Wide Web, ha traído consigo una serie de interrogantes en cuanto a su efecto sobre la práctica del periodismo y por tanto, sobre la aplicación de la documentación a estos nuevos contenidos periodísticos. Estas cuestiones se refieren básicamente a:

- Las variaciones en las características de los documentos periodísticos, tanto en su calidad como en su cantidad.
- Las nuevas propiedades de la documentación periodística derivadas de sus posibilidades de producción, de tratamiento documental y de utilización.
- La modificación en las relaciones del centro de documentación de prensa de un medio concreto con sus posibles usuarios: en primer lugar, con los periodistas, productores de los documentos primarios (los contenidos del medio), y al tiempo los principales usuarios del centro; pero también deben de tenerse en cuenta otros posibles usuarios: los demás sistemas del medio (el sistema de producción y, especialmente, el sistema de gestión del mismo, encargado de sostenerlo materialmente) y los usuarios externos.
- La nueva concepción económica del propio centro de documentación.

Todas estas cuestiones ejercen una gran influencia sobre la actualización de las tareas de los centros de documentación de aquellos medios que difunden sus informaciones, entre otros canales, a través de Internet. En esa actualización, como apuntamos al final del presente capítulo, el marcado de texto puede jugar un papel importante para la descripción de los contenidos periodísticos que permita al centro de documentación periodística adaptarse a la creciente evolución de los mismos.

3.2. Variaciones en las características de los documentos periodísticos

La consolidación del periodismo en línea ha planteado la cuestión sobre cómo este nuevo medio de difusión modifica las pautas de trabajo de los centros de documentación periodística.

A este respecto, y al igual que ocurrió con el surgimiento de otros medios de comunicación periodística respecto a la prensa escrita en papel, como la radio o la televisión, la difusión de información periodística a través de la web continúa cumpliendo las citadas características documentales básicas: enciclopedismo, universalidad, heterogeneidad, redundancia, dispersión, obsolescencia, ideologización y saturación.

Si bien estas características se mantienen, las diferencias comienzan a surgir en el grado en que se manifiestan. Factores como el enciclopedismo, la universalidad, la ideologización y la heterogeneidad se mantienen en los documentos periodísticos accesibles a través de la web en la misma medida que en cualquier otro soporte.

Sobre otro tipo de características, como la redundancia, la obsolescencia o la dispersión, el manejo de documentos digitales y su difusión en línea sí han supuesto una notable influencia, como se muestra a continuación.

3.2.1. Relación entre el aumento de la urgencia y las posibilidades de explicación

El factor tiempo es uno de los grandes condicionantes del trabajo periodístico. Los centros de documentación de los medios han debido responder a cualquier petición, habitualmente, en el menor intervalo posible. Los medios digitales han visto acrecentada la necesidad de que el lapso entre el momento en el que se produce un hecho y el momento en el que se difunde sea cada vez menor.

Ello se ha debido a que la cantidad y, por tanto, la frecuencia de información publicada no está en relación tan directa con el formato empleado: en los medios en papel, las necesidades de producción industrial obligaban a publicar medios con una frecuencia concreta y en un tamaño determinado, dados los costes de producción, aunque excepcionalmente esas restricciones pudieran ampliarse en momentos muy puntuales.

En los medios audiovisuales, aunque la información pueda ofrecerse de inmediato, los medios acostumbran a ofrecer compilaciones periódicas de información en aquellos momentos en los que se supone que las audiencias pueden ser mayores, así como cada cierta franja horaria (por ejemplo, cada hora), de modo que en muchos casos repiten la difusión de noticias previamente emitidas, ya sea tal cual o con ciertos detalles añadidos.

Esta misma secuencia se mantiene en los canales audiovisuales que dedican las 24 horas del día a ofrecer información continua, siguiendo el modelo de las agencias de información, pero ahora destinada a usuarios finales, y no solo a otros medios. Los lapsos entre espacios informativos con las noticias de actualidad se reservan para incluir reportajes específicos sobre temas concretos.

En el caso de los medios que difunden información a través de Internet, no hay necesariamente un periodo de cierre concreto, con lo que no existen las urgencias del cierre, o de la emisión. Pero, a cambio, las 24 horas del día pueden ser el momento del cierre; y especialmente, cuando se produce una información que va actualizándose continuamente. En el caso de la web, el medio puede *colgar* (es decir, poner accesible su información en un servidor) su información y dejarla ahí para que el usuario pueda acceder a ella en cualquier momento.

Además, al heredar la capacidad explicativa de los medios tradicionalmente impresos, estos medios digitales compiten también por ofrecer no solo una información, sino también, en lo posible, sus antecedentes.

Esa competencia hace que no solo sea el medio el que ofrezca más información a sus usuarios, sino que también, dentro del propio medio, haya aumentado el número y el

tipo de las peticiones al centro de documentación para complementar las informaciones: por un lado, los redactores pueden elaborar noticias relacionadas con un hecho concreto, y demandar información al centro de documentación para elaborarlas. Por otra parte, ha aumentado también el ámbito de los posibles demandantes: las peticiones pueden provenir, por ejemplo, de un corresponsal que se encuentre a miles de kilómetros, y que pueda acceder telemáticamente al centro de documentación.

¿Cómo debe reaccionar el centro de documentación? Aumentando el empleo de sus funciones previsorias y preparatorias, lo cual supondrá para ello mejorar las capacidades de identificación y descripción documentales, con el fin de tener un mayor control sobre sus fondos.

Dado que se le exige cada vez más información, más rápidamente y de mayor calidad, es preciso que su capacidad de análisis documental no solo se vea mejorada, sino que tienda a racionalizarse, tratando de reducir en lo posible el conjunto de tareas documentales meramente mecánicas para centrarse en plantear nuevas alternativas de difusión documental que permitan responder a esas necesidades incrementadas de sus usuarios. Pero además, debe tratar de conseguir, con ese mismo análisis, un mayor número de aplicaciones documentales concretas de ese esfuerzo. En otras palabras, que con un análisis documental puedan generarse tantos productos documentales tan rápidamente como sea posible, facilitando la generación automática de esos productos documentales.

Esa urgencia en la difusión periodística, y en el trabajo documental relacionado con ella, obliga al tiempo a considerar cuál es el grado de obsolescencia de los documentos recibidos. Al ser mucho mayor la capacidad de producción y de difusión de documentos periodísticos, una información puede ser considerada obsoleta en apenas unos minutos, sustituida por una nueva que actualice sus contenidos.

Pero, al mismo tiempo, esa información puede formar parte de un conjunto de documentos periodísticos que permitan rastrear la evolución de un suceso, mediante el establecimiento de relaciones adecuadas entre todos ellos, con el fin de que ese

conjunto documental pueda tener un valor periodístico, a corto o medio plazo, o incluso un valor futuro para investigadores en campos como la Historia, la Sociología, la Economía, etc.

Los documentalistas deberán de moverse, de forma acelerada en la continua consideración sobre lo que puede interesar en un futuro y lo que no, sobre lo que puede interesar más adelante o sobre lo que conviene desechar, sobre lo que puede ser redundante porque ya se tienen otros fondos similares dentro del medio... o no tanto, dado que convenga, por motivos ajenos a los periodísticos, conservar esos fondos. Y para ello, una ayuda en la toma de las decisiones puede ser la consideración de las relaciones establecidas expresamente entre las diferentes unidades documentales.

3.2.2. El riesgo de la saturación

Relacionado con los anteriores riesgos de obsolescencia y de redundancia en el conjunto de los fondos manejados por un centro de documentación, está otro enemigo en potencia: la saturación o alcance de la máxima capacidad de almacenamiento, que obliga al centro de documentación a eliminar aquellos documentos que no se consideran necesariamente útiles a corto, medio o largo plazo.

Todo centro de documentación debe mantener una política adecuada de conservación que evite llegar a niveles de saturación. Esa política deberá tener en cuenta, fundamentalmente, el principio de perdurabilidad –por razones obvias– y el principio limitativo, esto es, la capacidad material del propio centro de conservar los documentos que produce el medio. El empleo de documentos digitales ha permitido que, en un espacio mucho menor, la cantidad de posibles documentos a almacenar sea mucho mayor; pero esa capacidad, con ser relativamente barata, tiene un coste.

Por otra parte, debe considerarse un factor añadido de saturación: los medios periodísticos incluyen en muchos casos contenidos ajenos que son únicamente difundibles en un periodo concreto y para un ámbito concreto, más allá del cual el medio no puede reutilizar ese contenido, salvo para uso interno. ¿Hasta qué punto

conviene entonces almacenar esa información, si no es posible reutilizarla? ¿Debe entonces ser eliminada? ¿Conviene mantenerla con fines históricos, o de contextualización para contenidos relacionados en el futuro? ¿O es preferible, en su caso, eliminar el documento y mantener la referencia con vistas a que pueda ser de nuevo demandado al medio original; por ejemplo, la agencia o el agregador al que se compró el documento por primera vez?

Como en otros casos expresados anteriormente, se plantea una nueva disyuntiva. La primera opción, eliminar los contenidos que físicamente no puedan ser reutilizados – incluyendo aquellos que se consideren obsoletos, incluyendo también a los que han sido producidos por el medio– plantea la reducción del peligro de saturación, pero presenta a cambio el riesgo de pérdida absoluta de la información... en el caso de que el medio que originalmente produjo la noticia también considere obsoleto ese contenido, lo cual es en muchos casos difícil de precisar.

La segunda opción, conservar todo lo producido por el medio, ofrece la ventaja de mantener no solo la información, sino también los formatos en los que se produjo, con las ventajas ya descritas en el anterior epígrafe dedicado a la transformación. Empero, la cantidad de información producida por un medio periodístico alcanza volúmenes enormes incluso para el almacenamiento mediante soportes digitales, especialmente en aquellos medios que producen grandes masas de información continua, y que además ofrecen numerosas versiones de una misma información, conforme ésta se va actualizando.

A todo ello se une que, en la actualidad, cualquier medio periodístico digital puede producir versiones en varios formatos de sus contenidos, lo cual incrementa este peligro de saturación. En este caso, el principio limitativo es fundamental para determinar la decisión tomada en uno u otro sentido. Los medios pequeños disponen de menores recursos para mantener grandes archivos con su producción, mientras que los grandes medios generan enormes torrentes de documentos informativos multiplicados por los diferentes formatos.

La solución estará probablemente condicionada por el grado de utilidad futura que se le otorgue a los fondos por parte del medio. Por ejemplo, las agencias tenderán probablemente a no almacenar todas las posibles versiones de una noticia, dejando únicamente la versión final... con la dificultad de determinar cuál es esa versión final.

Los medios más pequeños habrán de asumir que, si mantienen sus contenidos en el futuro, esos fondos serán empleados como fuente de estudio histórico, geográfico, sociológico, político, etc. Pero, sobre todo, deben considerar que, con un pequeño esfuerzo de conservación, podrán probablemente reutilizar esos materiales y extraer de ellos un rendimiento económico: de hecho, como hemos apuntado, los primeros recursos de pago de los medios electrónicos fueron, y siguen siendo aún, sus “archivos” o hemerotecas digitales.

3.3. Nuevas respuestas a la dispersión documental

El fenómeno de la dispersión, entendida como la variedad de documentos, formatos, fechas y lugares en los que pueden aparecer documentos relacionados periodísticamente entre sí (por la materia, por los personajes que forman parte de la información, etc.) se mantiene también en los medios periodísticos accesibles a través de la web.

En este caso, no obstante, el uso de documentos digitales permite combatir en buena parte ese factor, mediante una serie de nuevas propiedades que requieren, eso sí, un notable aumento del control documental, tanto técnica como intelectualmente. Estos mecanismos son:

- a) La **estructuración**, o capacidad de los documentos digitales de ser divididos en tantas unidades como sea documentalmente necesario, con el fin de que esas partes puedan ser identificadas en cualquier momento, de modo que puedan aplicarse a sus contenidos las propiedades descritas a continuación, con el fin de permitir una mejor recuperación de los mismos.

- b) La **interrelación**, o capacidad de que los contenidos documentales del medio estén concebidos como un sistema en sí, en el que por una parte, al obtener una unidad documental se pueda también hacer referencia de modo cómodo y preciso a aquellos otros datos y unidades documentales que permitan contextualizar y, por tanto, hacer comprensibles los hechos y datos citados en el documento inicial.

Esta interrelación será posible aplicando diversos tipos de mecanismos de descripción a cada unidad documental. Uno de ellos será el uso de lenguajes documentales, sean lenguajes de clasificación jerárquicos o facetados; o bien, lenguajes de indización, como listas jerárquicas, listas de encabezamiento de materias, tesauros...; o cualquier otro tipo de representación, como ontologías, *topic maps*, etc. Otro de estos mecanismos, convenientemente enlazado con el anterior, será el establecimiento de relaciones mediante hiperenlaces, empleando para ello lenguajes de marcado como XLink y otros relacionados.

- c) La **reutilización**, ya sea de continente o de contenido. Entendemos por reutilización de contenido la posibilidad de que un documento o una parte del mismo pueda ser rápidamente empleado con un uso distinto del originalmente hecho (por ejemplo, un fragmento de una información que contenga declaraciones de una persona pueda ser empleado en otra información meses más tarde para contraponer esas declaraciones a otras del mismo personaje).

Entendemos por reutilización de continente la posibilidad de que ese documento o fragmento sea adaptable, dividible o añadido a versiones en otros formatos diferentes al empleado para producirlo (por ejemplo, que un documento en XML pueda convertirse en una versión PDF), con todo tipo de fines, ya sean documentales o de cualquier otro tipo.

Algunos de esos fines pueden ser: la inclusión en una base de datos que emplee diferentes lenguajes de almacenamiento y/o recuperación; la

reproducción a través de soportes que permitan solo determinados usos de determinadas partes, o viceversa, es decir, la unión de varios formatos para ser difundidos conjuntamente en un único soporte, como una página web; la transmisión a otros medios que usen formatos diferentes al empleado en el medio actual; la difusión de un mismo formato o de formatos diferentes en diferentes soportes, etc.

Ambos tipos de reutilización pueden ser simultáneos y, de hecho, la reutilización de continente implica una reutilización del contenido incluido. La variedad entre ambas reside, por tanto, en el empleo del mismo formato físico o en la generación de uno nuevo.

- d) La **integración**, entendida como la capacidad de los restantes sistemas del medio informativo (el sistema de producción periodística, el de control técnico de la producción y el de gestión económico-administrativa del medio, para asumir el uso del marcado de texto con fines documentales de modo que no solo no entorpezca el funcionamiento cotidiano de todos ellos, sino que, al contrario, mejore el rendimiento conjunto.
- e) La **multimedialidad**, o capacidad para que un mismo documento contenga fragmentos en diferentes formatos.

En los siguientes epígrafes se desarrollan con mayor profundidad estas nuevas propiedades.

3.3.1. Estructuración de contenidos: documentos, piezas, objetos y seudopiezas

¿Qué es hoy un documento periodístico digital? Partamos de las reflexiones sobre el concepto de documento digital para entender algunas de sus características básicas:

“Los documentos digitales son documentos, tienen mensaje, tienen soporte –aunque la unión de ambos no es ya indisoluble– y tienen

intencionalidad comunicativa (...) Esta característica que viene dada por la desvinculación contenido/soporte determina su virtualidad, y hace posible que los contenidos digitales puedan... ser accedidos remotamente, su transportabilidad. (...) El documento digital es además heterogéneo o complejo... puede encontrarse en varios archivos o servidores". (RODRÍGUEZ BRAVO, 2002, p. 479-480)

Además de poder concebirse como una realidad que puede estar físicamente contenida en varios soportes distintos, el documento digital posee otra capacidad: la de contener diferentes informaciones. Retomando la definición de **unidad documental**, citada en el capítulo 1, ésta se basa en una definición genérica apuntada por (GUINCHAT y MENO, 1992, p. 51):

La unidad documental es la parte de un documento que se ha aislado artificialmente con el fin de proceder a un tratamiento posterior: descripción bibliográfica, descripción del contenido, almacenamiento y búsqueda de información, difusión. Un documento constituye efectivamente una unidad física, pero puede contener informaciones de distinta naturaleza o sobre diversos temas, que puede ser interesante discriminar. (...) El único requisito es que la unidad sea físicamente reutilizable. (...)

Todo sistema documental debe establecer sus propias reglas con el fin de identificar las unidades documentales en función de la especificidad de los contenidos, del nivel de análisis deseado, de las necesidades de los usuarios y de las posibilidades prácticas. Las unidades documentales se tratan como si ellas mismas fuesen un documento, naturalmente con una remisión al documento original donde han sido identificados, y que, por su parte, constituye el objeto de un tratamiento propio.

¿Cómo puede aplicarse la unidad documental en la documentación periodística aplicada a contenidos digitales? Partamos de un caso concreto, alejado del periodismo, pero con analogías que se mostrarán a continuación: un libro de autoría múltiple, en el cual cada capítulo esté escrito por un autor distinto y trate de un tema diferente.

En ese caso, habría dos posibles consideraciones sobre lo que es una unidad documental: en primer lugar, desde un punto de vista intelectual, cada artículo podría ser considerado como una unidad documental, con significado propio. Y en segundo

lugar, el conjunto del libro podría ser considerado como una unidad documental, desde el punto de vista físico.

En el caso de los contenidos periodísticos, también podemos considerar cuáles pueden ser las unidades documentales desde un punto de vista físico y desde un punto de vista lógico, aunque las posibilidades. Tomemos en cuenta el siguiente caso:



Figura 3.1. Ejemplo de documento con dos piezas (una con dos objetos y otra con un solo objeto)

Fuente: *El País*, 27 de noviembre de 2003, p. 48

En este caso, el conjunto de los textos incluidos se refieren a un mismo acontecimiento: la respuesta de la Comisión Europea al acuerdo del Consejo Europeo de Ministros de Economía de no sancionar a Francia y Alemania en 2003 por incumplir el Pacto de Estabilidad al tener un déficit excesivo. Siguiendo, pues, el

razonamiento anterior, todo ese conjunto puede considerarse un **documento** (en este caso, una *noticia*).

Pero ese documento trata, desde el punto de vista intelectual, dos aspectos distintos: la propia respuesta de la Comisión Europea, por una parte, y una entrevista al exministro alemán de Economía, Theo Waigel, por otra. Cada uno de estos dos aspectos está recuadrado por separado en la figura anterior.

La división entre ambos aspectos tiene sentido porque, en un futuro, un usuario puede buscar únicamente uno de ellos, y no el otro. Consideraremos, por lo tanto, cada uno de esos aspectos intelectuales de un mismo documento como una unidad documental, a la cual denominaremos **pieza**. En la figura 3.1., la pieza 1 está formada por dos textos, y la pieza 2 está formada por un único texto.

Consideremos ahora la posible división física de esas partes. Debe observarse que cada uno de los tres textos pueden ser elaborados de distinto modo –por ejemplo, por diferentes autores–. Es decir, que, en origen, físicamente, pueden estar en tres ficheros físicos distintos. A cada una de esas unidades documentales físicas la denominaremos **objeto**.

Por lo tanto, una pieza puede contener más de un objeto, y un documento puede contener más de una pieza. No obstante, los textos periodísticos no siempre son tan fácilmente divisibles, al menos, en principio. En muchos casos, un documento periodístico está físicamente formado por un único objeto, pero incluye más de un acontecimiento: es lo que se conoce como **noticia compuesta**. Véase el siguiente ejemplo:

Sharon asegura que puede rebatir las acusaciones de corrupción **1**

Mueren tres palestinos, uno de ellos al intentar infiltrarse desde Siria

JULIO DE LA GUARDIA. Jerusalén. El primer ministro israelí, Ariel Sharon, aseguró poder probar su inocencia en el caso de supuesta aceptación de sobornos y financiación ilegal del partido que pende sobre él y sus hijos. Durante su visita al muro de separación de Cisjordania, Sharon se presentó como víctima de un complot políti-

co cuyo objetivo sería hacerle perder las próximas elecciones generales. Tres palestinos resultaron muertos en diferentes incidentes registrados a lo largo de la jornada de ayer, uno de ellos durante un intento de infiltración fronteriza desde los altos del Golán, que constituye el primero en la frontera con Siria desde que dióra comenzó la Intifada.

Acorralado por las acusaciones de corrupción y nepotismo que se inician de la investigación que está llevando a cabo la Fiscalía General del Estado, Ariel Sharon pasó al contraataque. Tras mantener tres días de silencio desde que se produjeron las primeras filtraciones a los medios de comunicación, aprovechando su visita a la primera fase del muro de 350 kilómetros que separará Israel de Cisjordania, declaró que "se trata de una confabulación política repugnante que voy a rebatir con documentos y datos". Según Sharon, este supuesto complot tendría como objetivo "tumbar al primer ministro".

En su vez, el coordinador de la campaña electoral del Likud y actual alcalde de Jerusalén, Ehud Olmert, reforzó la tesis de su candidato, añadiendo que "la mayoría tiene claro que esta historia es una burbuja". En cambio, la coordinadora de la campaña laborista y ex ministra de Inmigración, Yuli Tamir, presentó la interpretación contraria, asegurando que la información publicada hasta ahora es una muestra suficiente de la desho-

El amigo sudafricano

Por fin dio la cara también el ciudadano sudafricano que consiguió un préstamo de millón y medio de euros a los dos hijos de Sharon, Omri y Gilad, para que avalaran un crédito obtenido para sufragar los gastos que les supusieron las elecciones internas de septiembre de 1999, en que su padre se hizo con las riendas del partido. "Es amigo mío, y estoy muy contento de haber podido asistir a un amigo", declaró Cyril Kern sobre Sharon ante los micrófonos de la radio del Ejército. Asimismo, Kern negó durante la entrevista telefónica que hubiera tenido lugar cualquier tipo de conducta ilegal.

Por otra parte, en lo que supone un incidente de carácter racial, aunque se haya registrado algún precedente esporádico durante la Intifada, tres militantes fueron sorprendidos por

una patrulla fronteriza mientras intentaban infiltrarse en territorio israelí por la verja eléctrica que constituye la demarcación entre Jordania y Siria. Según la oficina del portavoz militar, el primero fue abatido por los soldados después de disparar contra ellos, mientras que el segundo fue herido y capturado, y el tercero consiguió huir en dirección a la parte siria de los Altos del Golán. Uno de los atacantes vestía un uniforme del Ejército sirio, aseguró la prensa fuerte.

Además, un taxista palestino de 33 años falleció al ser atropellado en el transcurso de un viaje en el campo de refugiados Yutuf, situado en el tercio sur de Jerusalén de Guza.

En otro incidente, un joven de 17 años fue abatido por disparos efectuados por los soldados al dispersar a un grupo de palestinos que se acercaron a presionar la demolición de la vivienda familiar de un supuesto terrorista de las Brigadas de los Mártires de Al Aqsa en la localidad de Solde, situada en la parte noroeste de Cisjordania.



Sharon habla a la prensa con el muro que Israel está construyendo de fondo. EFE

Figura 3.2.

Ejemplo de noticia compuesta

Fuente: *El País*, 9 de enero de 2003, página 6

En este documento se narran en realidad cuatro acontecimientos: la respuesta del primer ministro israelí, Ariel Sharon, a las acusaciones sobre corrupción que realizaron contra él ciudadanos de su propio país, en primer lugar; por otra parte, un suceso entre el ejército israelí y personas que intentaban infiltrarse en ese país; por último, dos incidentes entre palestinos y el ejército israelí, en diferentes ubicaciones, saldados en ambos casos con la muerte de ciudadanos palestinos. La noticia va acompañada de una fotografía en la que Sharon aparece hablando con la prensa, teniendo detrás a un hombre uniformado que parece estar situado a modo de protección.

Tomando en cuenta únicamente al texto de la figura (exceptuando el pie de foto), disponemos de un único objeto (documento físico). Pero, desde el punto de vista

intelectual, podría entenderse que una unidad documental sería cada uno de los *recuadros* marcados en la figura y señalados con un número (entendiendo que los conjuntos 2a y 2b forman parte de la misma unidad).

En este caso, denominaremos **seudopiezas** a cada una de las unidades documentales intelectuales físicamente incluidas en un mismo objeto.

En resumen, podemos distinguir distintos tipos de unidad documental que pueden estar relacionados entre sí:

- **Documento:** el conjunto lógico formado por una o varias piezas.
- **Pieza:** unidad documental que, a juicio del centro de documentación, puede ser recuperable, formada por uno o varios objetos, relacionados entre sí que pueden estar físicamente separados.
- **Objeto:** unidad física que contiene información y que puede ser relacionada con otras unidades, formando una pieza.
- **Seudopieza:** parte física insertada originalmente dentro de un objeto, la cual incluye un contenido distinto al del otras u otras partes del mismo objeto y que, por tanto, pueda ser susceptible de ser recuperada por separado por el centro documental. Un ejemplo: tomando la figura 3.2 como referencia, las cuatro divisiones realizadas corresponderían a sendas seudopiezas.

Pero la división podrá no ser necesariamente tan sencilla. El empleo de formatos digitales para representar los documentos periodísticos permite preguntarnos si sería posible elaborar nuevos documentos, empleando algunas combinaciones como las siguientes:

- Un documento podría estar formado por dos piezas, cada una de las cuales contendría objetos incluidos inicialmente en documentos distintos, y fusionados después en un nuevo documento por el centro de documentación.

- Un documento podría incluir versiones de un mismo objeto en formatos o lenguajes diferentes. Por ejemplo, se le puede dar a elegir al receptor de un medio multilingüe entre tres versiones en castellano, catalán y francés de un mismo contenido. En ese caso, dispondríamos de tres objetos... que solo varían en la lengua empleada. ¿Serían todos ellos una pieza? ¿Serían tres piezas diferentes (cada una de ellas, con su correspondiente objeto), dado que no hay otra relación entre ellas que el idioma, pero son el mismo contenido, y el receptor podría estar interesado en el texto en un idioma, pero no en los otros dos?
- Un nuevo documento podría contener los datos de una seudopieza incluida en un objeto de un documento, más una pieza incluida en otro documento distinto.

Estas y otras muchas posibilidades de combinación obligan a resolver otras cuestiones: ¿cómo relacionar cada una de estas divisiones? ¿Pueden reflejarse todas las posibilidades de combinación mediante una única definición de los contenidos que permita reutilizar los mismos según las necesidades de los usuarios? ¿Qué posibilidades ofrece esta nueva concepción para los centros de documentación? ¿Cómo afectan estas posibilidades al trabajo del centro de documentación? Las demás propiedades referidas a continuación se plantean como posibles respuestas.

3.3.2. Interrelación de documentos y componentes

El uso de documentos digitales incluye la capacidad de establecer entre ellos todo tipo de relaciones de modo que desde un documento se pueda no solo indicar la clase de relación que le une a otro –por ejemplo, temática–, sino acceder directamente, en su caso, a dicho documento relacionado.

El ejemplo general más conocido de esta capacidad de interrelación es el uso de hiperenlaces, como los que pueden emplearse en documentos HTML. La información periodística digital, si incorpora el hipertexto como modo de relación de

sus contenidos, permite establecer múltiples caminos de lectura, múltiples secuencias. (DÍAZ NOCI y MESO AYERDI, 1998, p. 5) hablan en realidad de la “ruptura de la secuencialidad” mediante el uso del hipertexto. Pero, a nuestro juicio, no se trata de una ruptura sino de una asunción distinta de este concepto; de ahí que no hayamos querido perder el término de “lectura secuencial”, dado que una lectura, es, por definición, una secuencia de atención sobre unos contenidos.

Por ejemplo, es posible relacionar, con apenas unos clics de ratón, un paquete de noticias publicadas sobre un tema concreto, si son convenientemente enlazadas, sin necesidad de volver a elaborarlas, como habría de hacerse en la versión en papel de un periódico, o bien, en un programa de televisión: la única forma de enlazar todas ellas, y de manera incompleta, sería elaborar un resumen. En el caso de un medio digital, basta con realizar una secuenciación adecuada de enlaces entre todas las noticias (por ejemplo, mostrarlas por orden cronológico), incluyendo además acceso directo a fuentes complementarias, materiales de diferente naturaleza (textos, vídeos, sonidos), etc.

Ya sea mediante hiperenlaces o mediante otros mecanismos de referencia entre documentos, los documentos digitales deben disponer de diversas capacidades para interrelacionar documentos como, sin ánimo de agotar las posibilidades:

- La búsqueda de contenidos comunes dentro de distintos documentos, por ejemplo, empleando como elemento de búsqueda la coincidencia de un texto en la misma parte de ambos objetos. Por ejemplo: titulares que contengan referencias a una empresa concreta.
- La elaboración automática de nuevos documentos, a modo de índices, basándose en los contenidos de ciertos elementos. Estos índices permiten el rastreo entre documentos, según tantos criterios como tipos de elementos puedan emplearse para elaborar los índices. Por ejemplo: la realización de índices sobre artículos basándose en sus títulos, sus autores, las fechas de publicación, los términos más citados en cada artículo, los nombres de los personajes citados en los artículos, etc.

- La relación de documentos elaborados en una fecha concreta con otros documentos elaborados previamente. Un ejemplo: la crónica del nombramiento de una persona para un cargo puede enlazarse con las entrevistas anteriores hechas a esa persona en ese medio.
- También puede suceder a la inversa: un trabajo inicial puede luego incluir enlaces con otros documentos posteriores que ayuden a explicarlo. Por ejemplo: una crónica sobre el estallido de un conflicto puede luego ir incorporando enlaces a una serie de artículos, noticias y reportajes que vayan explicando su desarrollo, en fechas sucesivas.

La capacidad de interrelación facilita, por tanto, el desarrollo de las funciones previsorias, preparatorias, crítico-verificadoras, completivas y rectificadoras. Esto es: no las modifica, ni añade nuevas funciones, pero sí las facilita y potencia. Con un uso adecuado, es más fácil elaborar productos documentales más completos y actualizados, reducir la necesidad de rectificaciones o prever, basándose en informaciones ya publicadas, las necesidades documentales de los periodistas ante nuevos acontecimientos.

La aplicación de mecanismos que interrelacionen las unidades documentales entre sí plantea numerosas preguntas: ¿cómo habrán de representarse, no solo periodísticamente, sino también documentalmente, los enlaces entre piezas, objetos y documentos que establezcan las relaciones entre los mismos? ¿Qué tipo de enlaces se utilizarán: los muy sencillos de HTML, o bien, enlaces con capacidades hipertextuales como los de XLink (ARCINIEGAS A., 2000)? ¿Qué otros tipos de relaciones, no solo hipertextuales, podrían establecerse entre estas unidades documentales? ¿Quién deberá hacerse cargo de la elaboración de esas relaciones? ¿Cuándo se establecerán las mismas: durante la elaboración de los documentos periodísticos o en un proceso posterior?

3.3.3. Reutilización de los contenidos documentales

Una de las grandes posibilidades que presentan los documentos digitales es la **reutilización**, es decir, que un documento pueda ser elaborado una sola vez, y que pueda ser empleado con múltiples usos, ya sea tal cual, por sí mismo, a petición de algún usuario, o modificado –convertido, ampliado, reducido, sustituido en todo o en parte, o eliminado– de modo que permita la elaboración de nuevos documentos a partir de otros ya elaborados. Y todo ello, sin que el contenido del documento se vea físicamente degradado.

Esta capacidad afecta no solo a los contenidos en sí –lo cual no sería un rasgo privativo de los documentos digitales: el índice de una publicación impresa puede ser una selección de los titulares de la misma, por ejemplo–, sino también a los contenidos, como acabamos de apuntar. Es decir, un objeto en un formato físico concreto puede ser fácilmente convertido a otro formato físico diferente –un texto en formato HTML puede pasar a ser un texto, por citar un caso, en formato Adobe Acrobat–.

Las posibilidades planteadas son, por lo tanto, enormes. Un mismo documento periodístico puede ser difundido, con una sola elaboración, de múltiples maneras:

- Mostrando solo los titulares, a través de herramientas que permitan el envío de mensajes cortos; el texto completo, a través de una versión en papel, o mediante copia digital de texto y formato, en un soporte electrónico.
- Ofreciendo el mismo texto del documento, con enlaces y referencias tanto dentro del mismo como desde otros documentos externos que enlacen esos documentos sin necesidad de modificar físicamente los documentos enlazados.
- Definiendo recopilaciones de informaciones o de extractos de informaciones, en nuevos documentos que puedan generarse de modo automático cada ciertos lapsos de tiempo...

Estas posibilidades ofrecen al centro de documentación todo tipo de oportunidades, pero también presenta algunos interrogantes. Los problemas afectan especialmente a los principios de adecuación funcional y de perdurabilidad, sin olvidar el limitativo: ¿qué es lo que tienen que almacenar? ¿Todas las versiones difundidas en distintos formatos de un mismo documento? ¿Sólo el documento original, completo, del cual se han extraído posteriormente fragmentos y/o se han realizado diferentes versiones formales para diferentes soportes?

La primera de las opciones estaría basada en la función modélico-narrativa del centro de documentación: si ciertos redactores le piden una determinada versión de un determinado documento, por ejemplo, con el fin de obtener ideas sobre cómo mostrar nuevos hechos en un nuevo documento basándose en enfoques anteriores, el centro debería poder responder a esa petición. Pero no solo en ese caso: también tendría sentido esta opción cuando los usuarios quisieran recuperar contenidos en un formato concreto.

No obstante, podría también optarse por una segunda alternativa de representación: un documento único que incluyera no solo el contenido completo de la noticia, sino también la información necesaria para disponer de la misma en distintos formatos, de modo que fuera posible reproducir la misma, en el formato deseado y en un momento concreto, tal y como ese documento, esa pieza o ese objeto fueron creados y reproducidos en su origen.

Para ello, haría falta disponer de la información sobre el contenido, y además, de todas las posibles estructuras y formatos que puede adoptar el documento (o alguna de sus partes) dentro del mismo.

La capacidad de transformación de documentos digitales ofrece al centro de documentación una oportunidad para potenciar su función informativa propia: el centro puede elaborar nuevos objetos, piezas e incluso documentos periodísticos completos que sean publicables por el medio, aprovechando la cantidad de información original almacenada dentro del centro de documentación y su conocimiento notable sobre dónde puede estar la información requerida, cómo

acceder a ella y cómo relacionarla.. En este caso, el papel del documentalista pasa a ser además el de periodista, en una integración de tareas cuyos requisitos y perspectivas de futuro están aún por descubrir.

También debe contarse con ciertos interrogantes desde un punto de vista documental, especialmente en cuanto a la capacidad de control de todos los documentos, piezas y objetos, y de las relaciones entre los mismos; y en cuanto a la capacidad de almacenamiento por el centro de documentación de todos los contenidos resultantes de la actividad periodística del medio, ya que el aumento de las posibilidades de difusión aumenta el riesgo de saturación. Ello obliga, a un enorme esfuerzo por parte del centro de documentación, para delimitar correctamente cada uno de los documentos y cada uno de los contenidos que debe controlar, con el fin de facilitar sus futuras utilidades.

Una posibilidad es el uso de herramientas informáticas como los sistemas de gestión de contenidos (SGC) (PALACIOS MARTÍNEZ, GARCÍA MORENO, y HERNÁNDEZ, 2002), que ofrecen ya implementaciones con las cuales, en principio, se permite un mayor control. Sin embargo, estos sistemas suelen requerir de ajustes para el flujo de trabajo de un medio concreto.

Los documentalistas deben participar en el diseño y mantenimiento de esos ajustes, como mínimo, para que la producción de los contenidos facilite el trabajo posterior de centro de documentación; por ejemplo, definiendo claramente en la fase de elaboración periodística de los contenidos la descripción de cada objeto y cada pieza, reduciendo el trabajo mecánico de los centros de documentación y facilitando la elaboración de trabajos más intelectuales, como la mejor definición de sus posibles usuarios y la elaboración de nuevas ofertas documentales para ellos.

3.3.4. Integración del centro documental con otros sistemas del medio

El sistema documental, que hemos denominado hasta ahora centro de documentación, forma parte de un entramado en el que también se incluyen otros

sistemas: el sistema editorial, el sistema de producción y el sistema de gestión. Dejando al margen este último, los tres sistemas son en realidad productores y receptores de información entre sí: la redacción y el sistema de producción envían sus documentos al centro de documentación. Pero también solicitan información al sistema documental, que les responde con esos mismos documentos, o con otros elaborados a partir de los primeros, ya sean resultados de una base de datos, informes completos que resumen datos publicados por el medio, etc.

En este sentido, los documentos enviados por los periodistas a los documentalistas están filtrados por el sistema de producción: los periodistas envían sus documentos al sistema de producción, que compone físicamente el medio, y los documentos finalmente generados son enviados también al centro de documentación.

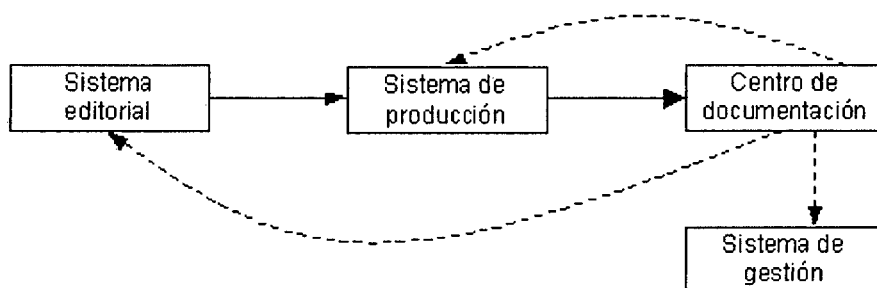


Figura 3.3.
Ciclo básico de los documentos producidos en un medio

Por lo tanto, deben conjugarse las necesidades de los periodistas –que usan unos formatos propios y tienen unas necesidades a la hora de elaborar los documentos periodísticos– con las necesidades del sistema de producción –que además de disponer de otros formatos para la difusión de los contenidos, debe poder controlar el estado físico de los mismos durante todo su proceso: cómo son, dónde están, en qué orden y posición deben utilizarse, cuáles son sus rasgos técnicos...– y, finalmente, con las propias necesidades del centro de documentación, que requiere la capacidad de describir lo más precisamente posible todos los contenidos recibidos con vistas a su posible recuperación.

La tendencia apunta a que los tres sistemas (y asimismo, el sistema de gestión) estén integrados. De este modo, los contenidos generados por el sistema de producción, por ejemplo, llegarían al centro de documentación en un formato que se adecuara, en lo posible, a las necesidades descriptivas del centro de documentación, facilitando al menos el reconocimiento automático tanto de las piezas y objetos utilizados como de ciertos contenidos fácilmente destacables: los titulares, la autoría, la ubicación del contenido dentro del medio, etc. Con ello, el centro documental ahorraría parte de la dedicación mecánica al análisis documental.

Para la integración de documentos periodísticos de modo que se adapten a las necesidades de cada sistema, se han desarrollado en la última década los citados sistemas de gestión de contenidos (SGC), herramientas informáticas que pretenden aglutinar al tiempo las necesidades redaccionales, productivas, y documentales (SYSIFUS, 2003).

Un SGC puede optar entre dos modos de integrar el funcionamiento de los distintos sistemas de un medio:

- Mediante diferentes formas de representación de contenidos para cada sistema, incluyendo herramientas que integren el formato empleado en otro sistema o que conviertan los documentos en un formato a documentos en un formato distinto. Este planteamiento choca con el hecho de que los formatos empleados en un sistema pueden variar mucho respecto a los formatos empleados en otros, con lo cual, de hecho, un sistema debería reelaborar los documentos recibidos para adaptarlos a su utilidad. Así puede ocurrir especialmente con el paso de los documentos del sistema de producción al centro de documentación.
- Mediante el uso, en lo posible, de un mismo formato de representación o, al menos, de formatos similares a que permitan una conversión relativamente fácil de uno a otro.

Esta segunda opción sería idealmente eficaz para el centro de documentación. Con ello, además, se facilitarían las posibilidades de reutilización de los contenidos.

3.3.5. Empleo de contenidos multimedia

Los contenidos informativos en línea poseen como característica básica la multimedialidad: en un mismo documento informativo (ya sea una noticia, un conjunto de noticias o una edición completa de un medio) pueden integrarse, de forma complementaria, textos, imágenes y sonidos, e incluso bases de datos o programas informáticos. Ello supone la posibilidad de un planteamiento radicalmente diferente sobre la elaboración, la recepción, la descripción, el almacenamiento y la recuperación de las informaciones.

En el caso de los medios audiovisuales, esta multiplicidad de soportes ya les era conocida: trabajar con los recortes de prensa exigía el uso de una serie de técnicas en parte distintas a las usadas para el trabajo documental con objetos periodísticos de sonido o de vídeo. Hoy, un medio digital puede manejar todos ellos y, por tanto, también tiene que almacenar y tratar todos esos soportes: no puede descartar información en ninguno de ellos porque, probablemente, cierta información no estará disponible más que en un soporte concreto, no siempre el textual.

A este respecto, (LAVIN, 2000) señala que “el audio y el vídeo pueden estar bien, pero no sustituirán la capacidad de la palabra para explicar”. No obstante, ciertos gráficos multimedia disponen de una mayor capacidad expresiva para la explicación de fenómenos concretos, desarrollados en un periodo temporal, que los respectivos gráficos estáticos tradicionales, como los mapas históricos.

Por poner un ejemplo, un mapa elaborado con tecnologías como Macromedia Flash, que permiten escenificar en un mismo gráfico el avance de unas tropas sobre un territorio durante un conflicto bélico, mostrando como van avanzando, puede resultar bastante más explicativo que su correspondiente explicación textual.

No obstante, la palabra siempre mantendrá su poder explicativo para razonamientos abstractos: de hecho, incluso los gráficos más expresivos incluyen texto como apoyo explicativo. Pero el texto no podrá ser un criterio excluyente del archivo de contenidos periodísticos en otros formatos: de ahí que, como ya hemos indicado, el análisis documental de los contenidos digitales multimedia requiera un tratamiento específico fuera del alcance de este trabajo.

3.4. Modificación de las relaciones del centro de documentación

La aplicación de las tecnologías digitales a los centros de documentación de prensa está provocando un profundo cambio en las relaciones del centro de documentación con todos sus posibles usuarios. Ese cambio se puede medir sobre todo en dos factores clave: la interactividad entre el centro y sus usuarios; y, relacionado con ese aspecto, la posibilidad de autoservicio ofrecida por el centro a los usuarios.

3.4.1. Aumento de la interactividad

Los centros de documentación de medios periodísticos han trabajado durante mucho tiempo al servicio principalmente de los redactores de los mismos. Pero además, han sido considerados en muchos casos como meros reductos a los que, muchas veces, no se les ha dado el valor oportuno.

La aparición de los soportes digitales puede cambiar radicalmente esta situación y, de hecho, ya lo está haciendo en algunos casos. Internet ha cambiado, de por sí, algunas de las prácticas informativas de los propios periodistas, en lo que a su relación con la documentación periodística se refiere. Baste el siguiente ejemplo, ya en 1997:

“The case of the Washington correspondent of a major national newspaper is illustrative. Before the Internet came along he would explain to the newsdesk in London what was the big story in the United States and write about it. Now London is searching the Internet and forever telling him what story to cover” (NICHOLAS y MARTIN, 1997, p. 49).

Esta muestra señala como muchos periodistas usan ya Internet como una herramienta documental a la que, en muchos casos, acceden directamente...

El centro de documentación dispone hoy de mayores capacidades de interacción con todos los posibles interesados en el manejo de sus contenidos. En primer lugar, con los periodistas: la comunicación, al menos en su aspecto telemático, dispone hoy de herramientas tanto de acceso directo –el ejemplo paradigmático es el correo electrónico, que agiliza la posibilidad de peticiones– como de oferta de contenidos desde el centro de documentación mediante instrumentos muy flexibles, como las Intranets, en las cuales el centro de documentación puede reforzar su función previsoría, adelantándose a las posibles peticiones de sus redactores, para ofrecer todo tipo de servicios que mejoren el trabajo de los periodistas: una selección de posibles fuentes de interés, la sugerencia sobre el uso de ciertos documentos, o el acceso mucho más cómodo a los contenidos ya publicados por el propio medio.

Pero, una vez hecho este esfuerzo de elaboración, se plantea que el uso de dichos recursos puede ser también accesible rápidamente para cualquier tipo de usuario potencial y no solo para la redacción. En otras palabras: una vez puesto un documento en su versión digital, ya sea en origen o convertida desde una versión analógica, ¿por qué no ofrecérselo a los gestores del medio para usos como el control de la publicidad ofrecida o el análisis de la cantidad de piezas publicadas por un periodista? ¿Por qué no ofrecerles a los usuarios el acceso cómodo a los contenidos ya publicados?

3.4.2. Autoservicio de los documentos por los usuarios

Uno de los grandes cambios que han surgido del uso de Internet entre los periodistas, los otros miembros del medio y los usuarios externos, consiste en que se han acostumbrado a buscar su propia información en la llamada “red de redes”: de una u otra forma, especialmente los periodistas –por estar acostumbrados a buscar información, aunque fuera de modos distintos a la búsqueda documental–, han comenzado a localizar información telemáticamente, mediante distintas

herramientas. La consecuencia derivada es que ahora demandan también poder buscar información directamente dentro del propio medio, de la misma forma que lo hacen fuera del mismo.

El papel del documentalista, por lo tanto, cambia en buena parte: deja de ser el custodio e intermediario de la información, en buena medida, dado que los periodistas no quieren pasar por un filtro: quieren acceder directamente a la información.

El problema, que probablemente ignoren en general los usuarios pero no los documentalistas, es que esa información inmediata y precisa no es tan fácil de conseguir: necesita previamente de un buen análisis documental y del diseño de herramientas documentales que permitan todo tipo de niveles de búsqueda, desde las consultas más sencillas hasta las más complejas, con el fin de que no se recurra expresamente al acceso directo a los documentalistas salvo en casos muy concretos.

Los documentalistas deben, pues, reforzar muchas de sus funciones con el fin de valorar las peticiones que les puedan ser hechas, *a priori*: en primer lugar, las clásicas funciones previsorias y preparatorias. Pero además, ahora deben ofrecer nuevas soluciones documentales accesibles a través de la red directamente por los posibles usuarios, internos o externos. Y deben ofrecer esas soluciones a los usuarios incluso antes de que estos se las puedan plantear.

Ello supone potenciar funciones como la informativa propia –los documentalistas pueden elaborar directamente ciertas informaciones, y no solo ofrecer datos a los redactores para que éstos compongan la información; por ejemplo, las recopilaciones de datos pueden ir ya redactadas para su publicación–, la modélico-narrativa –facilitando a los redactores ideas no solo sobre contenidos posibles, sino sobre formatos posibles a partir de los ya utilizados– o la orientadora, sin descuidar otras funciones más tradicionales como la función crítico-verificadora o la rectificadora.

Todas estas funciones deberán estar, por lo tanto, apoyadas en un ágil pero preciso análisis documental previo que permita tener en cuenta el desempeño de estas funciones en el futuro. Este análisis también debe permitir, como hemos señalado, la

rápida elaboración de cualquier tipo de productos documentales que sean fácilmente utilizables de modo independiente, en forma de **autoservicio**, por cualquier tipo de posibles usuarios.

Estos productos se están implementando ya, dando lugar, en el caso de los productos para usuarios externos, a los **servicios de valor añadido (SVA)**: las llamadas hemerotecas o archivos digitales son un caso claro de ello (FUENTES I PUJOL y GONZÁLEZ QUESADA, 1998).

No obstante, las posibilidades de recuperación, en estos casos, dejan aún mucho que desear, como puede deducirse, por ejemplo, de la muestra realizada por (JIMÉNEZ, GONZÁLEZ, y FUENTES I PUJOL, 2000), en un estudio que hoy sigue siendo válido. No se trata solo de poder almacenar los documentos digitales en bases de datos que permitan recuperar la información, sino de poder generar rutinas que asocien, manual o automáticamente documentos no relacionados inicialmente, basándose en factores comunes, reduciendo en cualquier caso el esfuerzo humano en realizar esas asociaciones.

Por ejemplo: al realizar una consulta en la base de datos de un periódico sobre las informaciones publicadas acerca de una película, podría encontrarse la ficha técnica de la misma el día de su estreno, pero también, tal vez la referencia a otra información sobre un premio que ha recibido una actriz por esa misma película, o el comentario que ha hecho un miembro del jurado que concedió ese premio, alabando la evolución de ese director desde su primera película, etc. La paradoja reside en que algunos de estos servicios de valor añadido ofrecen en la actualidad, con las herramientas utilizadas, un acceso mucho más limitado a todo ese contexto de información que las tradicionales carpetas incluidas en los archivos de recortes.

En resumen, el hecho de que los usuarios puedan acceder ahora a los contenidos documentales mediante el autoservicio obliga a los centros de documentación a tratar de hacerse visibles diseñando nuevos productos documentales más complejos, adelantándose a las necesidades de los usuarios y ofreciéndoles estos servicios, y permitiendo al tiempo obtener una rentabilidad económica para el medio, con el fin

de que los periodistas, los usuarios externos y también los gestores del medio contemplen la valía de sus esfuerzos (POITINGER, LARSON y otros, 2002).

3.4.3. Conservación: nuevos factores derivados de los documentos digitales

Los centros de documentación son también los encargados de conservar físicamente los contenidos producidos por el medio. En el caso de los documentos digitales, esta tarea se ve condicionada al menos por tres factores.

El primero de ellos es la capacidad de conservación de los documentos digitales. Si éstos no sufren físicamente desgaste, sí están amenazados a cambio un fenómeno tal vez peor: la **obsolescencia técnica**. En otras palabras, dada la rápida evolución de la informática, resulta difícilmente predecible que, a medio plazo, puedan recuperarse por completo los contenidos producidos hoy. Este problema, que puede deberse tanto a la falta de *software* (que el programa utilizado para generar ese contenido ya no sea utilizado hoy) como de *hardware* (que no se disponga de una herramienta física para leer el soporte en el que se encuentra el contenido).

Este problema provoca que los centros de documentación, no solo periodísticos (ROTHENBERG, 1995), tengan que contar con la posibilidad de enfrentarse a serios problemas para hacer accesibles a sus usuarios información generada en periodos relativamente recientes. Para paliar este riesgo, deberán plantear la representación de los documentos en soportes físicos (es decir, materiales) y lógicos (esto es, que estén hechos en lenguajes informáticos determinados) que ofrezcan a priori una mayor posibilidad de duración.

El segundo de los factores es la **pérdida de responsabilidad** sobre la custodia de los documentos. Por una parte, los documentos digitales forman parte de los contenidos gestionados por los ya citados sistemas de gestión de contenidos (SGC), que incluyen entre sus funciones la de almacenar y mantener todos los documentos generados en cualquier momento por cualquier miembro del medio. Ello ha supuesto que, en muchos casos, la responsabilidad última sobre la custodia de esos contenidos haya

pasado de los documentalistas –que físicamente disponían de los documentos en el centro de documentación– a los informáticos.

Probablemente, en consonancia con esta disminución, buena parte de los medios que comenzaron a publicar sus contenidos en Internet han sacrificado por completo sus contenidos publicados en los primeros años del fenómeno del periodismo digital, debido a motivos de “falta de espacio informático” y de “ahorro económico”, olvidando la función de “memoria histórica” del medio, con fines tanto periodísticos como de estudio para investigadores de diversas materias.

En la actualidad, paradójicamente, los medios han comenzado a tomar conciencia de que esos contenidos ya publicados pueden ser rentabilizados económicamente. De hecho, no solo se están manteniendo los contenidos completos de las publicaciones digitales, y su acceso al público, sino que también se está promoviendo el acceso digital a los fondos de periódicos publicados durante los siglos XIX y XX en papel. Un caso concreto es la hemeroteca digital del diario barcelonés (*LA VANGUARDIA*, 2003), que contiene todos sus números, en versión PDF, publicados desde 1881. No obstante, en este ejemplo concreto, la capacidad de reproducción del formato digital empleado deja aún que desear: las copias, especialmente de los ejemplares con más de 30 años de antigüedad, son difícilmente legibles.

El tercer factor, relacionado con el anterior, se refiere a la preservación de la **integridad** de los documentos. En la actualidad, con los debidos conocimientos, es relativamente fácil acceder a los sistemas informáticos que almacenan los contenidos de un medio periodístico. Hecho el acceso, es muy sencillo reproducir, sustraer o, peor aún, modificar los contenidos almacenados.

Esa preservación es una tarea asumida técnicamente por informáticos, que son quienes establecen físicamente las claves y los conocimientos para determinar cómo, dónde, cuándo y, en muchas ocasiones, por cuánto tiempo se almacenan físicamente, así como quién accede a ellos. Pero los documentalistas deben también jugar un papel importante: si no en el mantenimiento físico de esa integridad, sí en la definición de los principios que deben asumirse desde un punto de vista documental,

para mantenerla: con el empleo de las debidas aplicaciones informáticas, los documentalistas pueden realizar la administración cotidiana de esos fondos, aunque el mantenimiento técnico corra a cargo de informáticos. Los documentalistas, por tanto, tendrán que ampliar el conocimiento de herramientas informáticas, al menos, al nivel de usuario.

Sea cual sea el actor humano (informático o documentalista) que físicamente decida el acceso a los fondos documentales, éstos estarán probablemente integrados dentro del trabajo del medio a través de una red informática de modo que sea posible acceder a los contenidos tanto desde dentro del medio como, en su caso, desde fuera del mismo. Esa posibilidad plantea, a su vez, al menos dos interrogantes básicos: dentro del medio, ¿cómo controlar que un documento solo es accesible por quién debe tener permiso y potestad para ello? Y, fuera del medio, ¿cómo puede tener el usuario la certeza de que el respaldo del medio a ese documento que está viendo en su ordenador es fiable, y que el contenido que está visualizando no se trata de una variación colocada, por ejemplo, por un *cracker*?

3.5. Implicaciones económicas de los cambios en el centro de documentación

El acceso a través de Internet a casi cualquier medio periodístico existente desde cualquier lugar ha traído consigo también cambios para la documentación periodística desde un punto de vista económico. Esos cambios se refieren fundamentalmente a la independencia telemática respecto a los grandes distribuidores de la información, por parte del medio; la búsqueda de nuevos ingresos para los medios debidos al uso de sus archivos; y finalmente, la posibilidad de un mayor ahorro, en tiempo y dinero, para los usuarios de esos medios.

Hoy, se puede acceder a los contenidos completos de cualquier medio desde cualquier lugar del mundo, a un precio que –al menos en las zonas más avanzadas del planeta– puede considerarse óptimo en comparación con el precio de acceso a las

antiguas distribuidoras de información electrónica. Y, es más: ese acceso no tiene por qué realizarse a través de esos distribuidores, sino que los medios pueden comercializarlo independientemente, al precio que ellos asuman.

Ello no significa que no sigan existiendo intermediarios a esa información, incluyendo a aquellas empresas dedicadas a recopilar y organizar recortes de todo tipo de medios, aunque ahora sean en modo digital (AGUILERA CABALLERO, 2000), o que agrupen la mayoría de los contenidos de prensa de un país, como (My News online, 2003) en el caso español.

Este tipo de medios se han visto complementados por la aparición de una nueva vía de difusión periodística: los *agregadores*, o sitios web donde se recogen referencias de noticias que publican diversos medios periodísticos, con enlaces directos a esa noticia dentro del medio (FUENTES, JIMÉNEZ, y GONZÁLEZ, 2001) Algunos de estos servicios, en España, son (IMente, 2003), (Prensa Digital, 2003), Periodista Digital (ROJO, 2003), o, recientemente, la versión en español de Google News (GOOGLE, 2003).

Estos medios ofrecen un enlace a los contenidos concretos de las noticias publicadas por los medios, ya sea en el mismo día o, como en el caso de Google News, en los últimos treinta días. El interés de estos medios para realizar tales recopilaciones se basa en la inserción de publicidad, en la promoción de la tecnología empleada para su implantación futura en otros sitios web, o en la difusión del trabajo realizado para fomentar un mercado que pueda, pasado un tiempo, convertir el medio en un producto de pago.

En otros casos, son los grupos de prensa los que ofrecen la información producida conjuntamente por todos sus medios como valor añadido, directamente, sin intermediarios, aunque solo accesible para clientes de alto poder adquisitivo: empresas, instituciones, etc. Así ocurre con los casos de Factiva (REUTERS LTD., 2003), que agrupa los fondos de la agencia Reuters y de la editora estadounidense Dow Jones, la cual publica entre otros el diario económico *The Wall Street Journal*; o, en el caso español, Media Trader (VOCENTO, 2003), del grupo Vocento, que

agrupa los fondos de diarios como *El Correo*, *ABC*, una veintena de diarios regionales y otros medios de comunicación especializados.

Además de disponer de todos estos intermediarios, cualquier usuario puede acceder, al menos, físicamente, de modo directo a los contenidos de casi cualquier medio que exista; otra cosa es que se pueda acceder al contenido actualmente publicado por esos medios y a los contenidos retrospectivos, organizados en los respectivos archivos, o accesibles directamente mediante sendos enlaces.

La existencia de esos archivos, actualmente disponibles en muchos casos mediante una accesos de pago u ofertas económicas concretas, permiten ampliar el rango de posibles usuarios interesados especialmente en ejemplares concretos y, más aún, en documentos específicos, como un determinado artículo. La extensión de los posibles usuarios externos, sin embargo, es importante sobre todo a nivel espacial: cualquier persona puede demandar uno o varios contenidos, y recibirlos inmediatamente, sin retrasos ni posibilidad de pérdidas.

Al ampliarse el radio de posibles usuarios, también lo hace la posibilidad de que el medio pueda aumentar sus ingresos con los contenidos de sus archivos: no se trata ya de que solo puedan acceder usuarios corporativos, que puedan pagar una conexión durante largo tiempo a una base de datos cara o que puedan necesitar información importante por su rentabilidad económica, sino que cualquiera puede acceder por un precio relativamente módico a los contenidos ya publicados por un medio. ¿Es esta opción rentable, además de para el usuario, para el medio?

Algunos medios están empezando a plantearse alternativas de rentabilización de sus contenidos almacenados. En algunos casos, cuando el medio dispone de contenidos impresos que reproduce en un sitio web, opta por cobrar por el acceso a esos contenidos, como ocurre con *El Mundo*, en su sitio (elmundo.es, 2003), dejando que los contenidos publicados directamente en la web sean accesibles de modo gratuito en cualquier momento.

Otros medios, como *Expansión*, ofrecen en su sitio (Expansión.com, 2003) noticias de pago y noticias gratuitas, mezcladas. El hecho de reservar el acceso de pago a

ciertas noticias denota un análisis previo sobre la importancia de los contenidos ofrecidos. En el mismo sentido funciona el citado *La Vanguardia.es*, aunque en este caso, también se exige, para los contenidos gratuitos, que previamente el usuario se haya registrado en su base de datos. Finalmente, el modelo más restrictivo es ofrecido por (El País.es, 2003), que en su sitio web mencionado permite el acceso únicamente mediante pago, a todos sus contenidos excepto la página principal de su sitio web, continuamente actualizada, y a la sección de opinión.

Cabe citar que, en muchos casos, la disposición a cobrar por los contenidos incluye, como valor añadido, el acceso a todos los contenidos publicados por el medio desde su nacimiento. Los dos casos paradigmáticos en España pueden ser los del citado *El País*, que ha creado un acceso a todos sus contenidos digitalizados desde su nacimiento en 1977 en su versión accesible a través de la web; o, más aún, el mencionado *La Vanguardia*. Paradójicamente, en el caso de *El País* (aunque no sea el único ejemplo), no es posible acceder a los contenidos publicados exclusivamente en formato digital en los primeros años de su desarrollo.

Otros medios pequeños, más recientes, e incluso, únicamente existentes en su versión digital, también disponen de archivos con todos sus contenidos ya publicados, en modo gratuito, mientras que en algunas publicaciones, especialmente de periodicidad semanal, la restricción de acceso a los contenidos (o, sencillamente, la no disposición de los mismos) se restringe a los publicados en la versión papel y disponibles a la venta, dejando en algunos casos las versiones de archivo libremente accesibles, como ocurre con la revista (El Siglo de Europa, 2003).

En cualquier caso, los medios que ofrecen contenidos de pago no se remiten exclusivamente a la mera copia de sus contenidos impresos, sino que ofrecen otro tipo de servicios de valor añadido (SVA), cuya tipificación y utilidad documental ha sido desarrollada por (JIMÉNEZ LÓPEZ, 2001). Muchos de ellos precisan de una tarea documental previa para ser plenamente desarrollados: archivos de contenidos de cualquier tipo, envíos de información publicada basados en perfiles previos de usuario –alertas, difusión selectiva de información, etc.–, recopilación de información según temas concretos, convenientemente actualizados, etc.

La posibilidad de acceder telemáticamente a todos estos contenidos no solo es una fuente adicional para los medios, sino que permite un enorme ahorro para los usuarios, tanto en tiempo (especialmente para aquellos usuarios que se encuentran alejados del centro de producción del medio) como en dinero: resulta mucho más barato conseguir un ejemplar del medio a través de redes telemáticas que solicitarlo físicamente al medio y pagar además los costes del transporte.

Además, estos SVA exigen previamente en muchos casos que el usuario registre –a veces, incluso para el uso gratuito de esos servicios– ciertos datos, comenzando por su dirección de correo electrónico, con lo cual el medio consigue hacerse una medida de qué usuarios están realmente interesados a priori en sus contenidos, obteniendo una cartera inicial de posibles usuarios a los cuales ofrecerles cómodamente nuevos servicios de pago.

Por otra parte, el empleo del soporte digital como forma de producción, y de Internet como herramienta de difusión inmediata de sus contenidos, permite también que el usuario pueda decidir el modo en el que quiere recibir la información: una suscripción al medio, un ejemplar concreto, un documento específico dentro de un ejemplar, un servicio de valor añadido, etc.

Ello permite al usuario ahorrar su tiempo y su dinero en la búsqueda de información. Pero esta posibilidad de oferta supone también, a priori, un aumento de los recursos documentales destinados a organizar la información, por lo cual se requiere que se obtengan los mayores usuarios de un producto documental determinado, o bien, que el esfuerzo para realizar un producto documental determinado (por ejemplo, una búsqueda de un documento concreto, o la posibilidad de conseguir toda la información requerida sobre un tema) sea el menor posible.

En suma, la mayor y mejor capacidad de difusión de los contenidos documentales digitales del medio de comunicación a través de Internet han convertido a los mismos en un factor más a tener en cuenta a la hora de evaluar los posibles ingresos del mismo. No se trata, pues, solo de que el uso de esos contenidos documentales mejore el trabajo periodístico del medio, sino de la posibilidad de que, de modo directo, el

funcionamiento del centro de documentación pueda revertir en ingresos directos para el medio.

3.6. Interrogantes abiertos. El uso del marcado de texto

En el presente capítulo, se ha descrito la variación en las características documentales clásicas, modificadas por el uso de los documentos periodísticos digitales; se han mostrado las nuevas propiedades documentales derivadas de los mismos; se ha hecho hincapié en la modificación de las relaciones entre el centro de documentación y sus posibles usuarios; finalmente, se han mencionado algunos aspectos económicos derivados del aprovechamiento de las nuevas posibilidades de la documentación periodística.

El desarrollo de esos cuatro aspectos, en resumen, plantea algunas necesidades al centro de documentación periodística, relacionadas entre sí:

- ¿Cómo combatir el riesgo de redundancia documental y de obsolescencia de buena parte de los contenidos, que pueden ser rápidamente superados por nuevos contenidos publicados en poco tiempo por un mismo medio?
- ¿Cómo enfrentarse al tiempo a mayores y más complejas peticiones de información por parte de un mayor número de usuarios, en periodos cada vez más reducidos? ¿Cómo mejorar la capacidad de análisis documental reduciendo el lapso necesario para realizar ese análisis y facilitando su reaprovechamiento?
- ¿Cómo distinguir, con una única producción periodística, todas las posibles estructuras documentales que pueden derivarse de esa producción, y todas las posibles relaciones entre los documentos y/o sus diversas partes, con el fin de mejorar las capacidades de análisis documental y fomentar las múltiples posibilidades de reproducción?

- ¿Cómo integrar el funcionamiento del centro de documentación con los otros sistemas del medio –la redacción, los departamentos de producción y de gestión– de modo que los productos documentales se ofrezcan antes a todos ellos y, además, estén adaptados a las necesidades de estos sistemas, actuales y futuras?
- ¿Cómo conseguir aumentar el número de productos documentales para usuarios externos, de modo que éstos puedan acceder cada vez de un modo más fácil incluso sin ayuda directa del centro de documentación?
- ¿De qué modo preservar los contenidos del medio frente a riesgos como la posibilidad de saturación o la falta de control sobre los contenidos producidos y almacenados?
- ¿Cómo conseguir que los productos documentales y, por lo tanto, el propio centro de documentación, sean una fuente más de ingresos para el medio?

Trataremos de dar respuesta a estos interrogantes en los capítulos siguientes.

Referencias

El País.es. 2003. [Consulta: 16/10/2003]. <<http://www.elpais.es>>

El Siglo de Europa. 2003. [Consulta: 28/10/2003]. <<http://www.elsiglo-eu.com/>>

elmundo.es. 2003. [Consulta: 16/10/2003]. <<http://www.elmundo.es>>

Expansión.com. 2003. [Consulta: 16/10/2003]. <<http://www.expansiondirecto.com>>

IMente. 2003. [Consulta: 17/10/2003]. <<http://www.imente.com>>

My News online. 2003. [Consulta: 17/10/2003]. <<http://www.mynewsonline.com/>>

Prensa Digital. 2003. [Consulta: 17/10/2003]. <<http://www.prensadigital.com>>

AGUILERA CABALLERO, Begoña. Recortes de prensa digitales: todo un mundo de información a tu alcance. En: *VII Jornadas Españolas de Documentación. La Gestión del Conocimiento: retos y soluciones de los profesionales de la información*, Bilbao. Bilbao: FESABID - Universidad del País Vasco, 2000, p. 575-583.

ARCINIEGAS A., Fabio. What is XLink? En: *XML.com*, 2000. [Consulta: 20/7/2003]. <<http://www.xml.com/lpt/a/2000/09/xlink/index.html>>

DÍAZ NOCI, Javier y MESO AYERDI, Koldo. Desarrollo del periodismo electrónico. En: *El profesional de la información*, 1998, vol. 7, nº 12, p. 4-11.

FUENTES I PUJOL, M^a Eulàlia y GONZÁLEZ QUESADA, Alfonso. La prensa española en Internet. Análisis de los servicios de valor añadido. En: *Los sistemas de información al servicio de la sociedad. Actas de las VI Jornadas Españolas de Documentación*, 29 al 31 de octubre de 1998. Valencia: Fesabid, 1998, p. 281-292.

FUENTES, M. Eulàlia, JIMÉNEZ, M. Àngels, y GONZÁLEZ, Alfons. Servicios y sistemas de acceso a la información periodística en Internet. En: FUENTES I PUJOL, M^a Eulàlia (Dir.). *Anuario de Biblioteconomía, Documentación e Información. BIBLIODOC 2000*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 2001, p. 85-102.

GOOGLE. *Google News*. 2003. [Consulta: 17/10/2003]. <<http://news.google.es>>

GUINCHAT, Claire y MENOÛ, Michel. *Introducción general a las ciencias y técnicas de la información y documentación*. Madrid: CINDOC, UNESCO, 1992.

JIMÉNEZ, Àngels, GONZÁLEZ, Alfons, y FUENTES I PUJOL, M. Eulàlia. Las hemerotecas digitales de la prensa en Internet. En: *El profesional de la información*, 2000, vol. 9, n^o 5, p. 15-24.

JIMÉNEZ LÓPEZ, Àngels. *Estudi de la gestió documental de la informació en els serveis de valor afegit dels mitjans de comunicació a Internet: el cas de la premsa diària a l'estat espanyol*. Universidad Autònoma de Barcelona, 2001. [Consulta: 2/12/2003]. <<http://www.tdcat.cbuc.es/TDX-1008102-143437/index.html>>

LA VANGUARDIA. *Hemeroteca 1881-2003 (PDF)*. 2003. [Consulta: 15/10/2003]. <<http://www.lavanguardia.es/info/pdfhemeroteca.html>>

LAVIN, Rochelle Lewis. It's still journalism. En: *Poynter.org*, 2000. [Consulta: 9/4/2001]. <<http://www.poynter.org/centerpiece/082500.htm>>

NICHOLAS, David y MARTIN, Helen. Assessing information needs: a case study of journalists. En: *Aslib Proceedings*, 1997, vol. 49, n° 2, p. 43-52.

PALACIOS MARTÍNEZ, Marian, GARCÍA MORENO, M^a Antonia, y HERNÁNDEZ, Antonio. Gestión de contenidos en un centro de documentación de prensa. En: *Congreso Internacional de Información. INFO 2002*, La Habana. La Habana: IDICT, 2002.

POITINGER, Leigh, LARSON, Ron, JOHNSON, Nina, y EDDS, Carolyn. Skills team. En: *News libraries in crisis*, Minneapolis. Minneapolis: Special Libraries Association / Institute for New Media Studies-Minnesota Journalism Center, School of Journalism and Mass Communication, University of Minnesota, 2002, p. [Consulta: 2/3/2003]. <http://www.inms.umn.edu/convenings/newslibraryincrisis/information_team.html>

REUTERS, DOW JONES CO. *Factiva*. 2003. [Consulta: 16/10/2003]. <<http://www.factiva.com>>

RODRÍGUEZ BRAVO, Blanca. El documento: concepto y tratamiento. En: MORÁN SUAREZ, M^a Antonia y RODRÍGUEZ LÓPEZ, M^a del Carmen (coordinadoras.). *La documentación para la investigación. Homenaje a José Antonio Martín Fuertes*. León: Universidad de León, 2002, p. 467-486.

ROJO, David. *Periodista Digital.com. El periódico de periódicos*. 2003. [Consulta: 17/10/2003]. <<http://www.periodistadigital.com>>

ROTHENBERG, Jeff. Ensuring the longevity of digital documents. En: *Scientific American*, 1995, vol. 272, n° 1, p. 42-47. [Consulta: 12/11/2003]. <http://www.informatics-review.com/classic_reviews/long.pdf>

SYSIFUS. Gestión de contenidos desde cero. En: *Tierra de Nómadas*, 2003. [Consulta: 1/11/2003]. <<http://www.tierradenomadas.com/tw008.phtml>>

VOCENTO. *Media Trader*. 2003. [Consulta: 16/10/2003].
<<http://www.grupocorreomediatrader.com>>

Capítulo 4

LENGUAJES DE MARCADO PERIODÍSTICO: USO GENÉRICO PARA LA DEFINICIÓN DE SIGNIFICADOS Y DE ESTRUCTURAS DOCUMENTALES

4.1. El marcado de texto: síntesis histórica

Aunque el marcado de texto (o, en adelante, el marcado, a secas), es una práctica empleada desde hace siglos con fines tipobibliográficos, como explica (MARTÍN GALÁN, 2001, p. 232-235), el uso del marcado como método de representación documental aplicado a documentos digitales se remonta a finales de los años 60 y comienzos de los 70.

Hasta esa fecha, los distintos sistemas informáticos son incompatibles entre sí: resulta muy difícil intercambiar datos entre ellos, dado que los formatos de esos datos son muy distintos. Se plantea entonces la necesidad de una forma básica de representación en cualquier sistema que facilite ese intercambio entre diferentes sistemas informáticos y para múltiples aplicaciones.

Fue entre 1970 y 1973 cuando se desarrolló la primera de esas formas de representación, el Lenguaje de Marcado Generalizado, o GML. Este lenguaje era en realidad un *metalenguaje*: es decir, tenía como pretensión establecer las reglas sintácticas que debería cumplir a su vez cualquier otro lenguaje de marcado, ya fuera de propósito general, o bien, dedicado a la producción de documentos sobre alguna materia concreta.

El uso del marcado generalizado como mecanismo para representar la estructura de los documentos electrónicos comenzó a extenderse en los años 70, hasta el punto de que acabó convirtiéndose en una norma ISO, en 1986 (INTERNATIONAL STANDARD ORGANIZATION, 1986), con el nombre de SGML (Lenguaje Normalizado de Marcado Generalizado).

SGML ha servido para la elaboración de lenguajes muy especializados empleados en sectores complejos muy diversos –como la industria aeronáutica, la industria farmacéutica, el control de las sociedades bursátiles o la elaboración de textos humanísticos, por citar algunos (BOSAK, 1996)–. Pero fue un hecho mucho menos

trascendente, desde el punto de vista tecnológico, el que logró popularizar el uso de SGML: la difusión de un lenguaje basado en aquel, llamado HTML (RAGGETT, 1998). Éste último ha sido utilizado, desde 1995, para que millones de personas produjeran páginas web, empleando por tanto, sin saberlo, una gramática basada en SGML.

No obstante, HTML presentaba como inconveniente una capacidad muy limitada para expresar la estructura y el significado de las partes de los documentos. Inicialmente, surgieron algunas iniciativas para crear nuevos lenguajes más específicos, también basados en SGML, que permitieran crear estructuras más detalladas para difundir documentos a través de la web. Un ejemplo, en el ámbito periodístico, fue NIUTF (FIELD, 2001), desarrollo para noticias textuales, inspirado en HTML, pero añadiendo elementos propios de noticias periodísticas.

4.1.1. XML: de un metalenguaje a una familia de lenguajes de marcado

Pero la sintaxis de estos nuevos lenguajes basados en SGML era demasiado compleja y rígida para fomentar su uso masivo. Ello llevó al W3C (World Wide Consortium, la institución no oficial que, de facto, aprueba buena parte de los principales lenguajes empleados en la web) a plantear la posibilidad de generar un nuevo metalenguaje, mediante una versión más reducida de SGML, en la cual se pudieran basar nuevos lenguajes accesibles a través de la web.

Surgió así, en 1998, el Lenguaje de Marcado Extendido, más conocido por sus siglas XML, un lenguaje basado en un subconjunto de SGML. XML, en palabras de sus creadores, tenía como objetivos

“to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML” (BRAY; SPERBERG-McQUEEN y otros, 1998).

La paradoja reside en que la mayoría de los documentos accesibles en la web siguen estando basados en HTML: al menos, en la versión final que puede visualizar un usuario en su navegador, hasta el punto de que incluso se ha definido una versión de HTML basada en XML, con el nombre de XHTML (PEMBERTON, 2000).

XML, sin embargo, ha conseguido en apenas cinco años ser una referencia ineludible para la generación de lenguajes de marcado, al permitir que puedan definirse nuevos lenguajes con estructuras y elementos adaptados a las necesidades de cada colectivo (por ejemplo, para la difusión de documentos periodísticos), sin necesidad de que todos los documentos generados con un mismo lenguaje compartan la misma estructura fija. En este sentido, XML ofrece la misma capacidad de SGML, pero con unas reglas mucho más sencillas.

En apenas cinco años, han surgido lenguajes propios, basados en XML, con elementos diseñados expresamente para designar conceptos y partes propias en documentos de áreas muy especializadas, empleados en muchos casos en todo tipo de ámbitos especializados, a veces alejados de la propia web.

Más aún, muchos lenguajes basados en SGML se han convertido, con las oportunas actualizaciones, en lenguajes basados en XML; siguiendo con el mencionado NIUTF, en 1996 se convirtió en NITF, y en 1999, se redefinió siguiendo las reglas de XML (KARBEN, 2002).

Por otra parte, la definición de XML como un metalenguaje destinado a la creación de nuevos lenguajes de marcado desde un punto de vista documental ha sufrido un vuelco hasta convertirse, hoy, en un lenguaje destinado también a la definición de datos muy estructurados susceptibles de ser almacenados en grandes bases de datos para su procesamiento posterior. Esta doble posibilidad de descripción ha provocado el surgimiento de una creciente industria de desarrollo y mantenimiento de lenguajes basados en XML, si bien, al precio de aumentar la complejidad del propio metalenguaje.

Hoy, XML se ha convertido en una familia de lenguajes, basados en la gramática XML, con propósitos cada vez más específicos: por citar algunos, la definición de

enlaces hipertextuales, mediante el lenguaje XLink (DEROSE; MALER, y ORCHARD, 2001), o la definición de metadatos o datos descriptivos sobre un documento, mediante el lenguaje RDF (LASSILA y SWICK, 1999), por citar algunos ejemplos.

4.2. Descripción de estructuras y significados mediante marcado de texto

Siguiendo la descripción dada por (COOMBS, RENEAR, y DE ROSE, 1987), pueden distinguirse hasta seis posibles técnicas del marcado para describir diversos aspectos de un documento: el *marcado de puntuación*, que ofrece información sintáctica sobre expresiones escritas; el *marcado de presentación*, que permite mejorar estéticamente la presentación de textos; el *marcado de procedimiento*, que consiste en el uso de instrucciones informáticas (comandos) dentro de un programa para que sistematice pautas de presentación en un trabajo (por ejemplo, que le aplique un mismo estilo a todos los títulos de un trabajo); el *marcado referencial*, que se emplea para hacer referencia a entidades externas a un documento; el *marcado descriptivo*, que permite identificar las partes estructurales de un trabajo; y finalmente, el *metamarcado*, o metalenguaje de marcado de texto, que permite definir una serie de reglas básicas para elaborar otros lenguajes de marcado descriptivo.

En el presente trabajo, nos centraremos en estos dos últimos usos, el **marcado descriptivo** y el **metamarcado**, con la pretensión de identificar las distintas partes de que se compone un texto. En particular, consideraremos dos tipos de marcado distintos, basados en la concepción anterior:

- El **marcado estructural**, que consiste básicamente en la identificación de las posibles partes de que consta un documento, su orden de aparición, su disposición dentro del mismo y el establecimiento de las posibles combinaciones que puedan realizarse entre dichas partes. Tales fracciones

del documento se mantienen independientemente del modo en que se reproduzcan, visual, auditiva o táctilmente. Por ejemplo: una división de una noticia en titular, subtítulo, entradilla y texto (a su vez, desglosado en párrafos), puede mantenerse, independientemente de que la misma estructura pudiera ser mostrada de diferentes maneras: cambiando los tipos de letra para cada parte, mostrándose con colores diferentes según el soporte en el que se muestre la noticia, etc. Esta estructura, no obstante, puede incluir diversos grados de flexibilidad, aunque todos los elementos disponibles y todas las posibles combinaciones entre los mismos deben estar previamente especificadas.

- El **marcado semántico**, que consiste en la identificación, basada en un documento, de sus contenidos, de sus cualidades y del entorno en el que esos contenidos tienen unos significados, según múltiples grados de profundidad.

Entendemos en este caso por **contenidos** el conjunto de información que puede ser aprehendida expresamente por un receptor al acceder a dicho documento. Por ejemplo, al leer una noticia en un periódico impreso, un lector puede acceder al texto completo de la misma (y, en su caso, a la fotografía), incluyendo además el nombre del autor de la noticia, la ciudad desde la que se redactó y otros datos que puedan incluirse expresamente en el texto (por ejemplo, el nombre del medio en el que se originó la noticia, si se publica en un medio distinto).

Entendemos por **cualidades** de un documento el conjunto de datos asociados al documento que no forman parte del contenido del mismo, pero que sí permiten su descripción documental con vistas a su mejor recuperación. Algunas de ellas, siguiendo con el ejemplo periodístico, pueden ser el nombre del medio en el que se publica, la fecha de publicación, la sección en la que se publica, la página, el tamaño del documento, su ubicación física (por ejemplo, si se dispone de una versión en un soporte electrónico accesible en línea), los derechos de propiedad sobre el contenido, etc. Esos contenidos estarían incluidos en lo que, en el ámbito de la información electrónica, se conoce como *metadatos*.

Finalmente, entendemos por **entorno** el conjunto de información no incluida expresamente en el contenido pero necesaria para la comprensión del documento; parte de ella es añadida al documento en su tratamiento posterior por un centro de documentación, con el fin de aclarar su significado: bien del documento concreto, bien de algunas de sus partes estructurales, bien de alguno de sus fragmentos físicos (en el caso del texto, palabras o frases determinadas) que, contempladas en el conjunto del documento, pudieran requerir una explicación suplementaria. Un ejemplo puede ser la mención al nombre de una persona como referencia al cargo que ocupa, que un receptor puede no conocer en un futuro: Jaume Matas, por “Ministro de Medio Ambiente (2000-2003)”, pero por “Presidente de Baleares (2003-)”.

4.2.1. Algunas disyuntivas del análisis de contenido: autores y modos de descripción

Tradicionalmente, los centros de documentación periodística, en lo que a documentos periodísticos se refiere, han denominado **catalogación** al proceso de descripción de las cualidades de un documento periodístico, y **análisis documental** o **de contenido** al proceso de descripción de dicho contenido o del entorno en que se enmarca.

El análisis documental supone la representación del contenido o de su entorno mediante términos, tomados de un sistema de clasificación o incluidos en un proceso de indización, ya sea en lenguaje libre o empleando un lenguaje controlado. Estos términos pueden ser asignados por personas o bien ser aplicados a los documentos por programas específicos que aplican diversos algoritmos informáticos sobre la ocurrencia de palabras y expresiones en los textos.

En cualquier caso, las operaciones documentales de descripción y recuperación de documentos se basan en lo que hemos denominado el entorno de los mismos, aunque en este caso el entorno tenga un valor cuantificable mediante un programa informático, o bien, cualitativo, dependiente de la interpretación humana.

En el caso automático, este entorno se obtiene asumiendo, por ejemplo, que el uso continuado de ciertas palabras en ese y en otros documentos similares indica que el entorno definido por esas palabras es *relevante* a la hora de identificar el documento; en el caso de la asignación de términos por personas, se asume que el clasificador o indizador identifica el documento mediante términos que relacionan ese documento con otros similares, tratando, en lo posible, de anticipar los términos con los que un usuario puede tratar de localizar, ahora y en el futuro, ese documento.

Esta postura, que permitiría economizar el enorme esfuerzo, humano y económico, sobre todo de la indización manual de documentos periodísticos, ha chocado con diversos problemas. Como indican Hernández y Caridad, además de los problemas económicos de la indización, deben de plantearse

“los problemas lingüísticos, puesto que la temática y el sentido de un texto no lo otorgan solo el conjunto de términos que componen el texto sino también su contexto, sus sinónimos y otras figuras lingüísticas como las paráfrasis, las ambigüedades intencionadas, etc (...)

... en el caso de la documentación periodística, de carácter universal y enciclopédico, los resultados de la indexación automática no son tan halagüeños puesto que la representación de los textos exclusivamente por los términos presentes en el mismo, que es lo que hasta ahora hace la indexación automática, es claramente insuficiente puesto que no resuelve los múltiples problemas lingüísticos del análisis de contenido.”
(HERNÁNDEZ PÉREZ y CARIDAD SEBASTIÁN, 2000, p. 52)

Esta reflexión teórica se apoya además en datos cuantitativos concretos:

“In my dissertation research (Weinberg, 1981, p. 80 [(WEINBERG, 1983)]), I found that approximately 10% of humanly assigned index terms do not occur in full text. This has been confirmed in subsequent studies which demonstrate that controlled vocabulary indexing enhances full text retrieval by 10% (Hersh & Hickam, 1995) [(HERSH y HICKAM 1995, p. 746)]” (WEINBERG, 1996).

Y además, se complementa con la visión pragmática, desde un punto de vista periodístico y técnico, por parte de los responsables de algunas de las mayores bases de datos de prensa:

“Information users will be able to find documents that have been accurately and consistently indexed from a controlled vocabulary. This is true whether working with print or electronic information resources and is particularly relevant in today’s wired environment. Controlled language access always facilitates efficient information retrieval.”
(SYKES, 2001, p. 2)

4.2.2. Relación entre la descripción semántica y la descripción estructural: modelos de marcado

En cualquiera de las combinaciones entre el autor (ya sean personas o programas informáticos quienes describan los documentos) y el modo en el que se realice el análisis documental (mediante lenguaje libre o lenguaje controlado), el uso del marcado de texto como herramienta documental ofrece una utilidad fundamental para mejorar ese análisis: la capacidad para relacionar la descripción semántica de un documento y la descripción estructural del mismo.

Es decir, con el marcado de texto, además de describir semánticamente los contenidos, características y entorno de cada documento, se puede describir también la estructura de cada documento, por distinta que sea. Pero también se puede aprovechar la definición de esa estructura, la distinción de las partes del documento, para asignar una descripción semántica a cada una de esas partes: en este caso, a cada documento, pieza, objeto o seudopieza, e incluso a fragmentos muy pequeños dentro del texto, si es preciso.

El marcado semántico, por lo tanto, puede ser considerado también desde el punto de vista de la estructura de un documento, es decir, sobre a qué parte o partes de la estructura de un documento se le aplica. Visto así, pueden utilizarse tres posibilidades o modelos para la asignación física de la información semántica respecto a la estructura del texto que se pretende describir: el **marcado semántico inserto** junto al propio texto, el **marcado semántico adjunto** y, finalmente, el **marcado semántico externo** al documento en el que se encuentra el texto.

En el caso del **marcado semántico inserto** dentro del propio texto, aunque en la visualización del documento no se muestre, la descripción del entorno está unida al propio texto al que se refiere. Por ejemplo, el texto

Raúl sufre una rotura de ligamento

exige una explicación añadida. Si en el texto se sigue leyendo

...del jugador de la selección española de fútbol...

se entiende que el nombre de ese jugador está acotado. No se puede tratar, por ejemplo, de otro jugador, también llamado “Raúl”, que no haya jugado nunca en la selección española de fútbol, siguiendo con el ejemplo. Sin embargo, la constatación de la identidad exacta del jugador no es aún completa. Suponiendo que se refieran a una noticia aparecida el 7 de junio de 2003, y revisando la lista de jugadores del partido disputado por la citada selección, surgen dos jugadores que se denominan “Raúl”. ¿Cuál de ellos es? Una posible solución sería representar ese fragmento del documento mediante el marcado, tal que así:

```
<noticia>
  <personaje nombre="Raúl González Blanco">
    Raúl</personaje> sufre una <lesion>rotura de
    ligamento</lesion>
</noticia>
```

Figura 4.1.
Ejemplo básico de noticia con marcado contextual inserto

Como se verá más adelante al hablar de lenguajes de marcado concretos basados en este modelo, las posibilidades de descripción contextual inserta son muy amplias y pueden llegar a un gran nivel de complejidad... que también plantea interrogantes sobre su uso.

Con el nombre de **marcado semántico adjunto al documento**, denominamos a aquel que forma físicamente parte del documento como tal pero que no está mezclado con el contenido documental, sino que se coloca al principio o al final del

mismo. Siguiendo con el ejemplo anterior, un posible marcado adjunto, basado en XML, sería:

```
<noticia>
  <identificacion>
    <personaje nombre="Raúl González Blanco"/>
    <suceso tipo="lesion" detalle="rotura de
      ligamento"/>
  </identificacion>
  <contenido>
    Raúl sufre una rotura de ligamento...
  </contenido>
</noticia>
```

Figura 4.2.
Ejemplo de marcado contextual adjunto

Por último, en el **marcado semántico externo al documento**, la información sobre el contexto de un determinado documento está situado en un documento aparte, de forma que ambos están relacionados mediante, por ejemplo, una clave, un enlace, una combinación de ambos, etc. En el siguiente ejemplo, el documento llamado *noticia.xml* contiene el citado texto

```
<noticia>
  Raúl sufre una rotura de ligamento
</noticia>
```

Figura 4.3.
Ejemplo de noticia básica, sin descripción contextual

Y el documento *descripcion.xml* contiene la información contextual sobre *noticia.xml*:

```
<descripcion>
  <documento direccion="noticia.xml"/>
  <personaje nombre="Raúl González Blanco" />
  <suceso tipo="lesion" detalle=" rotura de
    ligamento" />
  <momento fecha="7-6-2003" />
</descripcion>
```

Figura 4.4.
Ejemplo de descripción externa de una noticia

La elección de uno u otro modelo de marcado conlleva una serie de implicaciones, tanto documentales como de otros tipos, que señalamos a continuación.

4.2.3. Implicaciones de los modelos de marcado

¿Por qué plantearse el uso de uno u otro de estos modelos? ¿Para qué tipo de descripción documental? ¿Supone el uso de uno u otro de los modelos la asunción de un enfoque documental diferente? ¿Qué otros condicionantes podrían tenerse en cuenta a la hora de decidir la aplicación de un modelo concreto?

El uso de uno u otro modelo de marcado debe tener en cuenta una serie de factores, desde un punto de vista documental:

1. **La determinación sobre cuál será la unidad documental empleada en el centro de documentación.** No es lo mismo que el centro esté acostumbrado a trabajar con documentos cortos que puedan ser en sí una única unidad documental formada por una única pieza; o bien, que el documento pueda tener múltiples estructuras diferentes, y que exista un gran porcentaje de documentos complejos.
2. **El tipo de representación adoptado para el análisis documental.** También es importante plantear si se pretende que las descripciones documentales se realicen mediante lenguaje libre, lo cual evita la definición de lenguajes documentales controlados, o bien si se optará por éstos últimos, con lo que se tendrá que calibrar el modo más fácil de uso de esos lenguajes tanto en su creación y mantenimiento como en su aplicación para el análisis concreto.
3. **La consideración de las unidades documentales como elementos modificables o no.** Partimos de la consideración de que el contenido de la unidad documental no puede ser modificado, puesto que es competencia de su autor.

Pero sí es posible determinar si el análisis documental se realizará en documentos secundarios físicamente distintos de las unidades documentales analizadas, con lo cual físicamente no habrán de modificarse esas unidades, o bien, si los documentalistas podrán añadir descripción documental posteriormente, insertada dentro del contenido o adjuntada al mismo.

4. **La posibilidad de conexión de las unidades documentales y de su correspondiente análisis documental.** Si el medio dispone de posibilidades para conectar de modo fácil y fiable el fichero físico que contiene la unidad documental con el fichero físico, distinto, que contiene el análisis documental, podrá optar por esta opción. Si, por el contrario, resulta complicado o costoso o poco fiable a largo plazo mantener físicamente separados contenido y análisis, habrá de plantear la posibilidad de que la unidad documental contenga físicamente la descripción adjunta o añadida.
5. **Las facilidades e inconvenientes físicos ofrecidos por cada modelo para la recuperación de información.** Este aspecto incluirá la consideración sobre cómo será de fácil o de complejo el acceso físico a las unidades documentales y a sus descripciones por parte de las herramientas informáticas de recuperación que se quieran emplear.

Partiendo de estos cinco interrogantes, el **marcado inserto** permite distinguir todas las posibles unidades documentales que puedan definirse, por pequeñas o grandes que sean, facilitando además la distinción de la estructura que esas unidades documentales formen dentro del documento. Además, el marcado interno permite detallar la descripción de cualquier detalle noticiable dentro de cada unidad. Dicho de otra forma: permite que puedan asignarse descriptores específicos para partes muy concretas, ya sea un párrafo o una mención expresa a un personaje, por pequeña que sea su participación en el hecho.

El uso de este modelo de marcado, aunque no impide la aplicación de un sistema de clasificación, es más proclive al empleo de descripción documental mediante indización, ya sea manual o automatizada, al permitir la elaboración de un gran número de puntos de acceso para la recuperación del documento. También es más proclive a la utilización de lenguaje libre, dado que basta con señalar ciertos puntos de un documento para convertirlos en posibles entradas de un índice generado automáticamente.

Esta posibilidad puede ser especialmente útil a la hora de describir documentalmente una *noticia compuesta*: sin necesidad de alterar el texto de la misma, puede dividirse su contenido total en varias piezas informativas mediante el oportuno marcado de texto.

A cambio, el marcado inserto presupone que, al menos documentalmente, las unidades documentales y los objetos físicos son modificables, dado que resulta difícil que durante la elaboración de un documento periodístico el autor emplee el marcado para señalar todos los posibles puntos susceptibles de describir documentalmente su texto: esta tarea deberá ser realizada en todo o, al menos, completada posteriormente, por el centro de documentación.

Ello implicará un nuevo acceso y modificación física de los ficheros correspondientes que contengan el documento, si no del contenido del mismo, sí del marcado empleado. Si los autores modifican posteriormente el documento, también habrá de revisarse de nuevo el marcado realizado por los documentalistas, lo cual puede hacer muy complejo el proceso.

Un marcado inserto supondrá que, en caso de que una pieza pueda estar formada por dos objetos descritos por separado, puedan existir discrepancias en el marcado empleado: diferentes tipos de profundidad (por ejemplo, que en un caso se identifiquen todos los personajes, y en el otro, no); diferentes descriptores, incluso contrapuestos (por ejemplo, “acuerdo parcial”, en un caso, y “desacuerdo relativo”, en otro, refiriéndose al mismo acontecimiento) o falta de enlace entre ambos documentos, por error u omisión en la referencia a ambos documentos en el sistema documental.

El marcado inserto, además, permite el empleo de lenguaje libre, incluyendo el uso de herramientas que marquen automáticamente la aparición de ciertos términos y empleen cálculos relacionados con la frecuencia de los mismos para obtener posibles valores sobre la relevancia de los mismos a la hora de ser utilizados para recuperar el documento en un futuro.

Ello no significa que este tipo de marcado impida el empleo de términos descriptivos, no incluidos expresamente en el texto, mediante lenguajes controlados –por ejemplo, como *atributos*–. Pero, al tiempo, dificulta la búsqueda de esos nuevos términos, al estar las descripciones mezcladas con el texto del contenido. Una posible solución es el empleo de herramientas documentales como los ficheros invertidos, aplicables a los términos empleados en las descripciones, a los contenidos o a ambas posibilidades.

Además, el marcado inserto puede suponer un notable aumento físico del tamaño de un documento, que dependerá de la cantidad de marcado añadido al contenido original. Ese aumento puede suponer que físicamente la cantidad de información a manejar pueda desbordar las posibilidades físicas de las herramientas informáticas (programas y máquinas) disponibles en el centro de documentación.

El **marcado adjunto** también posibilita que los documentos sobre los que se apliquen sean modificables, por los mismos motivos citados al hablar del marcado inserto. Esta opción permite también el uso de lenguaje libre o controlado, aunque reduce la posibilidad de indexación automática basada en la ocurrencia de términos, al menos, empleando el propio marcado como mecanismo de localización de los términos.

A cambio, facilita el acceso automático a las correspondientes descripciones de las cualidades y del entorno de cada documento, dado que la descripción, aunque físicamente esté dentro del documento, está claramente distinguida de ésta al comienzo o al final del mismo. Esta circunstancia puede reducir el tiempo de búsqueda y acceso a los documentos si se utiliza un programa o robot destinado a la recuperación de documentos basándose en esas descripciones.

Un marcado adjunto permite además agrupar, por ejemplo, la descripción de varios documentos lógicos bajo un mismo documento físico (un único fichero), lo cual permite compartir descriptores documentales, lo cual también ayudaría, desde un punto físico, a reducir el nivel de saturación del centro documental.

Pero, a cambio, este modelo de marcado aumenta el riesgo de producir ruido documental. Retomando un ejemplo concreto: una información sobre la lesión del futbolista Raúl González, en la que el parte médico de la lesión formara parte de un documento físico que contenga al tiempo la crónica del partido, una fotografía y un despiece con declaraciones de los futbolistas... y solo al final, el citado parte médico. En este caso, si el descriptor documental sobre la lesión estuviera al comienzo del documento, provocaría un retardo innecesario hasta que el lector encontrara, leyendo el documento, la información concreta sobre el percance médico.

Por último, el **marcado externo** plantea mayores dificultades para la descripción de unidades documentales menores que el mero objeto, dado que, si fuera así, sería difícil indicar dónde empieza y dónde termina cada una de las unidades documentales (o seudopiezas) distintas incluidas dentro de un mismo objeto.

Además, el uso del modelo externo presenta la misma posibilidad de generar ruido documental que el marcado adjunto. Por lo tanto, al igual que éste, sí puede ser útil para el marcado sobre cualidades (catalogación) pero disminuye el uso de la estructura del documento como factor para la indexación: es decir, no permite emplear la división en partes del documento para añadir a cada una de ellas un análisis por separado.

Este modelo presenta también la facilidad de que una herramienta automática de búsqueda actúe exclusivamente sobre los ficheros que contienen las descripciones realizadas durante el análisis documental, lo cual también podría reducir el tiempo de acceso físico a esas descripciones.

Pero, frente a esa ventaja física de recuperación, el marcado externo requiere que el documento original no sea modificado posteriormente, que sea un documento cerrado para que una modificación posterior del original no difiera de su descripción. Y además, el hecho de no estar físicamente unido al original supone el peligro de que no sea posible acceder desde la descripción del documento al contenido del mismo, ya sea por causas físicas (por ejemplo, un error en la red de conexión entre ambos) como por causas conceptuales (por ejemplo, una referencia desde la descripción al

documento original en la cual la dirección del original no esté correctamente escrita... o haya cambiado sin que haya advertencia de ello).

En resumen, el siguiente cuadro comparativo muestra las diferencias básicas entre los tres modelos de marcado apuntados:

Marcado inserto	Marcado adjunto	Marcado externo
Mezcla contenido con descripción	Separa intelectualmente el contenido de su descripción	Separa físicamente el contenido de su descripción.
Permite distinguir y describir todas las partes del documento (incluso las más pequeñas)	Obliga a emplear una única descripción para varios objetos o piezas	Dificulta la distinción de partes de un documento
Permite el uso de documentos modificables, aunque plantea problemas entre la elaboración y la descripción.	Permite el uso de documentos modificables	Dificulta el uso de documentos modificables
Facilita la indexación automática mediante el mero reconocimiento de contenidos marcados, pero también permite el uso de indización manual	Permite el uso de indexación automática y de indización manual	Permite el uso de indexación automática y de indización manual
Dificulta el uso de buscadores sobre las descripciones documentales	Reduce las dificultades para el uso de buscadores sobre las descripciones documentales	Facilita el uso de los buscadores sobre las descripciones documentales
Aumenta el tamaño físico de los documentos	Aumenta el tamaño físico de los documentos, en menor medida que el marcado inserto	Separa en distintos ficheros el contenido de los documentos de su descripción documental, lo que reduce el tamaño físico requerido por fichero.

Marcado inserto	Marcado adjunto	Marcado externo
Riesgo: ruido documental por exceso de descripción	Riesgo: ruido documental por excesiva distancia en el documento entre descripción y contenido	Riesgo: silencio documental por no poder localizar físicamente el original referido en una descripción.

Figura 4.5.
Tabla-resumen de tipos de marcado

La discusión sobre el empleo de uno u otro modelo no queda agotada aquí; especialmente, porque los distintos lenguajes de marcado ya definidos para la descripción de contenidos periodísticos susceptibles de ser empleados con fines documentales emplean en su mayoría modelos híbridos de representación que no responden totalmente a ninguna de las opciones planteadas. A la descripción de estos lenguajes específicos y de sus características de uso documental, por separado, dedicamos el resto del presente capítulo.

4.3. Evolución histórica del mercado de textos periodísticos

Para poder comprender los diferentes lenguajes de marcado periodístico actualmente propuestos, debemos previamente establecer una breve sinopsis histórica que muestre su génesis y su desarrollo actual.

A principios de los años 70, comenzó la introducción en algunos medios periodísticos de los ordenadores como herramientas de trabajo, principalmente, en Estados Unidos. Ello trajo pronto consigo la pretensión de conseguir que, utilizando las primeras redes telemáticas, se aprovecharan las noticias generadas por agencias (que, incluso hoy, siguen siendo buena parte de los contenidos difundidos por cualquier medio periodístico), transmitiéndolas directamente desde los ordenadores de las agencias hasta el sistema informático de los medios que recibían sus noticias,

con el fin de que fueran aprovechadas directamente, en lugar de ser recibidas en papel a través de teletipos y reintroducidas a mano en el sistema informático del medio por los redactores.

Con este fin, comenzaron a trabajar en el tema instituciones como ANPA, la Asociación de Editores de Periódicos de Estados Unidos, rebautizada luego como NAA; y el consorcio IPTC, siglas del Consejo Internacional de Telecomunicaciones de Prensa, formado fundamentalmente por la mayoría de las grandes agencias de prensa de todo el mundo –Reuters, AP, UPI, AFP, Efe, etc.– y por algunos de los mayores editores de periódicos del mundo, como *The New York Times* o Dow Jones, la editora de *The Wall Street Journal*, entre otros. A estos medios iniciales se han unido posteriormente empresas dedicadas a la redistribución de contenidos periodísticos ajenos a través de múltiples medios: es lo que se denomina hoy *agregadores* o, en una terminología previa a la irrupción de Internet como forma de difusión periodística, *sindicadores de contenidos*.

Las primeras iniciativas (FIELD, 2001) se dirigieron hacia la creación de lenguajes de marcado que permitieran la transmisión de noticias textuales entre agencias y otros medios, añadiendo para cada noticia, además del texto, algunos otros datos, ya fueran de descripción básica, ya fueran de carácter técnico (inicio y fin de transmisión, formato físico del texto, etc.). Así surgió, en 1974, el primer lenguaje de este tipo en EEUU, el ANPA-1170, modificado por el ANPA-1228, en 1976. Éste último acabó por ser adoptado internacionalmente en 1979, tras su aprobación por el IPTC como IPTC TEC-7901. Con diversas variaciones, este lenguaje fue utilizado para la transmisión telemática de noticias entre medios durante los años 80, de modo que acabaron por ser dos los principales lenguajes utilizados, a finales de los 80 y comienzos de los 90, para la transmisión telemática de noticias: el ANPA 1312, mayoritario en Estados Unidos, Canadá, América Latina y Asia, y el IPTC 7901, mayoritario en Europa. (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003b)

A comienzos de los 90, en el seno del IPTC comenzó a plantearse, por una parte, la necesidad de transmisión de contenidos periodísticos no solo textuales, sino también

icónicos, sonoros y audiovisuales. Por otra parte, también se planteó, en vista de que las herramientas informáticas y las redes telemáticas eran cada vez más potentes, la posibilidad de que las descripciones de las informaciones periodísticas, muy básicas hasta entonces, tuvieran mayor profundidad. A ello contribuyó el, si cabe, mayor aumento en la cantidad de informaciones que podían ser transmitidas, lo que suponía una mayor complejidad en el tratamiento de las mismas dentro de los medios. En el seno de los mismos, comenzaron a instalarse los primeros **sistemas de gestión de contenidos**: es el caso, por ejemplo, del diario madrileño *El Mundo* (ARENAL, 1995).

A estos dos requerimientos, transmisión de formatos no solo textuales y descripción más profunda de los contenidos, se unió el desarrollo de Internet como herramienta de transmisión periodística. Los tres factores determinaron el desarrollo de los lenguajes de marcado periodístico.

En el caso de la transmisión de contenidos, surgieron entre 1991 y 1997, los llamados por el IPTC **modelos de intercambio de información**, más conocidos por sus siglas, IIM (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003b) que pretendían la creación de un lenguaje mediante el cual se pudiera transmitir cualquier tipo de información periodística independientemente de su formato, y establecer además diversos campos de descripción del contenido. IIM estaba basado en el modelo físico OSI (Interconexión de Sistemas Abiertos), aprobado en 1984 (INTERNATIONAL STANDARD ORGANIZATION, 1994) y desarrollado en una serie de normas ISO utilizadas de modo habitual en la transmisión telemática de contenidos durante los años 90. Por tanto, IIM se basaba en la descripción de un documento en *capas* de elementos, que contenían en su mayoría detalles técnicos de transmisión y de descripción del formato, así como algunos descriptores temáticos con el fin de ser usados para clasificar cada información periodística, además de una capa que encapsulaba el contenido completo del documento.

IIM fue desarrollado, pero solo en parte, mediante varios documentos que fueron siendo publicados sucesivamente: una descripción genérica del lenguaje y su

implementación futura, denominada *Anatomy of a wire story 11* (FIELD; DOHERTY y otros, 1992), en septiembre de 1992: una primera versión (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL y Newspaper Association of America, 1993), en abril de 1993, que definía aspectos técnicos sobre el uso del lenguaje en ordenadores Apple Macintosh; una *Guía de Implementación* (HALL, 1994) que explicaba el encapsulamiento de textos en formatos IPTC 7901 y ANPA 1312, en abril de 1994; una nueva *Guía de Implementación* (ROCHE, 1997) que definía un conjunto de códigos temáticos o *Subject Codes* de propósito general, usados para la descripción de las noticias; y, finalmente, una versión, en octubre de 1997 (IPTC-NAA, 1997), que recoge todos los anteriores documentos descritos hasta el momento, y remite a otras clasificaciones como DNPR (Archivo de parámetros de fotos periodísticas digitales) (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL y Newspaper Association of America, 1997), que especifica entre otros detalles los campos de descripción textuales que deben acompañar a una imagen para su transmisión entre distintos medios.

Enlazando con el objetivo anterior, la transmisión telemática de noticias, IPTC había comenzado a desarrollar, por las mismas fechas, los mencionados **códigos temáticos** (*Subject Codes*) (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003d), unas clasificaciones por materias cuyo fin era facilitar el uso y control documental de las informaciones por los propios medios (véase un ejemplo en el **Anexo 6** de este trabajo).

Estas clasificaciones estaban formadas por una serie de listas de autoridades con varios niveles jerárquicos de descriptores, tanto alfabéticos como numéricos, con los cuales se pretendía elaborar una descripción temática más precisa que la empleada hasta entonces. Todo ello perseguía generalizar los términos empleados para la identificación de cada documento, para facilitar su intercambio entre diferentes medios, y su manejo dentro de un medio concreto.

Las categorías eran inicialmente de propósito general y pretendían clasificar cualquier tipo de noticias, aunque inicialmente se caracterizaran por dos tendencias: la descripción más profunda de aquellos temas sobre los que se publicaban mayor

cantidad de contenidos periodísticos (por ejemplo, la economía, los deportes o los sucesos), y, además, el sesgo anglosajón en las clasificaciones. Pronto, sin embargo, el carácter internacional de IPTC hizo que estas clasificaciones fueran poco a poco traducidas a otros idiomas, como el español (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003c). IPTC ha continuado el desarrollo de estos códigos desde entonces, fomentando su inclusión otros lenguajes definidos por este consorcio, como se explica más adelante.

Finalmente, relacionadas con las anteriores iniciativas, surgió la necesidad de plantear lenguajes de marcado que permitieran la elaboración de noticias transmisibles a través de Internet. La primera iniciativa en este sentido fue el ya citado lenguaje NIUTF (Formato de Texto Universal para la Industria Periodística) en 1995 (CROWLEY, 1995), un desarrollo basado en SGML y muy influenciado en particular por HTML: muchos de sus elementos utilizaban la misma denominación. La explosión de Internet como herramienta de comunicación periodística a partir de 1995 corrió pareja al surgimiento de XML como metalenguaje de marcado en 1998, lo que provocó la rápida adaptación de NIUTF, renombrado en 1996 como NITF, a una versión basada en XML en 1999 (FIELD, 2001).

A partir de 1998, comenzaron a desarrollarse lenguajes concretos con pretensiones ya de ser aplicables de modo masivo. Dos fueron las grandes tendencias en este sentido: por un lado, las iniciativas del citado IPTC para elaborar lenguajes que permitieran combinar la descripción semántica y la descripción estructural de los documentos periodísticos. Por otra, la pretensión de emplear lenguajes más volcados en la descripción semántica que en la definición estructural. Estos últimos estaban basados en iniciativas más genéricas de descripción mediante metadatos, como RDF. Una tercera opción, inicialmente descartada, fue la combinación de ambas posibilidades.



4.3.1. Lenguajes de marcado semántico basados en metadatos

La opción menos compleja, desde el punto de vista técnico, para la descripción semántica de documentos periodísticos consiste en el mero empleo de metadatos, incluidos generalmente en documentos independientes al empleado para contener sus metadatos. Los dos lenguajes que, fundamentalmente, han desarrollado esta tendencia, son RSS y PRISM.

RSS fue definido originalmente en 1999 por Netscape (PILGRIM, 2002) como un formato para construir portales que ofrecieran titulares tomados de sitios periodísticos que ofrecían noticias completas; a estos sitios periodísticos se les denominó *canales*. Aunque Netscape abandonó pronto esta pretensión y, por ello, el lenguaje, éste fue aprovechado por otros pequeños productores y distribuidores de contenidos, no solo sitios web periodísticos, para la difusión de metadatos sencillos sobre sus contenidos, con el fin de fomentar la ya descrita sindicación: en este caso, la difusión de sus contenidos a través de otros sitios web.

RSS presentaba la ventaja de su gran simplicidad –lo que le hacía un lenguaje de fácil manejo, especialmente para pequeños productores de contenido–, pero ha sufrido un enfrentamiento frontal entre sus valedores, que ha llevado a la escisión en dos versiones, la 1.0 y la 2.0, que se obvian entre sí, como se comenta más adelante.

La otra gran iniciativa es PRISM (IDEALLIANCE, 2003), conjunto de lenguajes desarrollado por un consorcio de editores de publicaciones periodísticas y científicas, así como de otros productos editoriales (libros, etc.); dicho consorcio está liderado por la empresa IdeAlliance, dedicada al fomento de especificaciones basadas en XML.

PRISM se basaba originalmente en la definición de metadatos que describían documentos, especialmente impresos, con el fin de facilitar su intercambio y reutilización. Para ello, se tomó como referencia RDF, empleando además algunos mecanismos de otros lenguajes anteriores también dedicados a la descripción de contenidos periodísticos, como XMLNews (surgido en 1999 y luego descartado por

la industria periodística, aunque algunos de sus principios sí han sido incluidos en lenguajes posteriores).

PRISM presenta una mayor profundidad de descripción documental que RSS (incluso en la versión 2.0 de éste último, con un mayor número de elementos descriptivos).

4.3.2. Lenguajes definidos por IPTC: el uso de marcado semántico-estructural

La segunda tendencia en el uso de marcado de texto para contenidos periodísticos, como hemos apuntado, tendía a la definición de lenguajes que no solo sirvieran para describir semánticamente los contenidos incluidos en un documento, sino que también permitieran determinar la estructura documental, con el fin de permitir el tratamiento de cada una de las partes.

Cronológicamente, como hemos apuntado, la primera iniciativa en este sentido fue el lenguaje NITF (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2001), surgido en su versión actual en 1999. NITF, no obstante, es un lenguaje destinado al manejo, estructural y semántico pero solo de noticias textuales, aunque incluyera algunos elementos para referenciar objetos informativos de audio y vídeo, pero a un nivel básico.

A pesar, por tanto, del desarrollo de NITF en sucesivas versiones (hasta llegar a la 3.2 en el año 2003), se hacía necesaria la creación de nuevos lenguajes que pudieran transmitir y describir información en cualquier tipo de formato, tal y como ya se había comenzado a intentar con IIM.

Dos fueron las propuestas en este sentido, desarrolladas entre 1999 y 2000. La primera de ellas fue la definición de XMLNews (MEGGINSON, 1999), que incluía a su vez dos partes: un desarrollo modificado de NITF 2.5 (la versión entonces vigente); y, además, un lenguaje para la descripción de metadatos aplicados a

cualquier tipo de contenido periodístico, ya fuera texto, imagen, sonido o multimedia, basándose en el citado RDF.

La segunda propuesta, que acabaría triunfando, fue presentada por la agencia Reuters, miembro del consorcio IPTC, en el seno del mismo, con el nombre de NewsML (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL, 2003). Como más adelante se explica en profundidad, NewsML permite transmitir, simultáneamente, diferentes objetos periodísticos, agrupados o no por piezas (por ejemplo, diferentes versiones de un mismo texto periodístico en formatos físicos distintos: NITF, PDF, Word, etc.), añadiéndoles además metadatos para cada objeto, para cada pieza o para todo un documento NewsML en conjunto. NewsML también está basado en XML.

Esta posibilidad de transmisión e identificación a medida –cuyo espíritu estaba ya presente aunque no totalmente desarrollado en IIM–, unida a la posibilidad de especificar la estructura (al menos, al nivel mínimo de objetos) dentro de documentos periodísticos complejos, hicieron que NewsML fuera rápidamente aceptado en el seno de IPTC y promocionado por sus miembros, muchos de los cuales poseen una notable fuerza en el seno de la comunidad periodística internacional. NewsML ha sufrido ligeras variaciones desde su primera versión, en 2000, hasta la actual en vigor (la 1.2., en octubre de 2003).

Al tiempo, IPTC continúa desarrollando lenguajes de marcado aún más especializados que NITF, para el tratamiento de noticias en campos concretos; en el caso de los deportes, IPTC ha hecho público el lenguaje SportsML ((INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003a)), con una versión aprobada en mayo de 2003. Además, está desarrollando un lenguaje adicional, ProgramGuideML (THE JAPAN NEWSPAPER PUBLISHERS & EDITORS ASSOCIATION (NSK) e International Press Telecommunications Council, 2003), destinado a la difusión de información textual sobre programas de radio y televisión, aunque aún se mantenga en fase de borrador.

4.3.3. El futuro de las iniciativas de mercado periodístico: entre la competencia y la convergencia

PRISM y RDF están más próximos al concepto de *web semántica* desarrollado por el W3C, en cuanto a que siguen el planteamiento seguido por este consorcio para definir lenguajes, también basados en RDF, que permitan crear colecciones de documentos de metadatos externos a los documentos originales.

Esta concepción se enfrenta inicialmente a la idea, desarrollada por el IPTC, de asociar estos metadatos a la estructura de los documentos, de modo que, necesariamente, los autores de los documentos y los de los metadatos sean los mismos, al menos, desde el punto de vista de la institución en la que se produzcan los unos y los otros.

Pese a esta percepción inicial, las utilidades en uno u otro sentido no están del todo claras. Como más adelante se explica en profundidad, en el empleo de NewsML puede convivir la creación de ciertas descripciones estructurales de documentos con la asignación de metadatos externos, debido a un criterio básico de NewsML (y también, de otros lenguajes desarrollados por el IPTC): la modularidad.

Por otro lado, la parte no normativa de la declaración de PRISM asume la existencia de NITF y de NewsML, la coincidencia en ciertos aspectos y la necesidad de establecer mecanismos que permitan la interrelación entre PRISM y los lenguajes del consorcio IPTC, lo cual se ha plasmado en el llamado PAM (PRISM Aggregator Message), que será comentado más adelante.

Consideramos oportuno, por lo tanto, realizar una descripción, desde el punto de vista documental, de cada uno de estos lenguajes, a la que seguirá el planteamiento, en un capítulo posterior, de las capacidades comparadas y combinadas sobre el uso de todos ellos, con el fin de mejorar el tratamiento documental de los contenidos periodísticos.

Referencias

ARENAL, Luis. "El Mundo": documentación integrada en la producción diaria. En: *Boletín de la SEDIC*, 1995, n° 21, p. 3.

BOSAK, Jon. An SGML-Web Server. En: *Fifth International World Wide Web Conference*, Paris. Paris: 1996. [Consulta: 2/4/2003]. <http://www5conf.inria.fr/fich_html/slides/dday/sgml/slide3.htm>

BRAY, Tim; SPERBERG-McQUEEN; C.M., y PAOLI, Jean. *Extensible Markup Language (XML) 1.0*. 1998. [Consulta: 30/6/2003]. <<http://www.w3.org/TR/1998/REC-xml-19980210>>

COOMBS, James H, RENEAR, Allen H., y DE ROSE, Steven J. Markup systems and the future of Scholarly Text Processing. En: *Communications of the ACM*, 1987, vol. 30, n° 11, p. 933-947. [Consulta: 4/5/2003]. <<http://www.oasis-open.org/cover/coombs.html>>

CROWLEY, M. J. *Marking text using NIUTF*. 1995. [Consulta: 28/4/2003]. <<http://www.intnet.net/pub/JOURNALISM/NIUTF.markup.language.info>>

DeROSE, Steve; MALER, Eve, y ORCHARD, David. *XML Linking Language (XLink) Version 1.0*. 2001. [Consulta: 30/6/2003]. <<http://www.w3.org/TR/xlink/>>

FIELD, Jeff; DOHERTY, Peter; MOSHER, Sue, y ALLEN, David. *Anatomy of a wire story*. 1992. <http://www.iptc.org/download/iimv4.zip>, fichero: Anatomy2u.pdf>

FIELD, Jeffrey. *IT standards in the news. The history of international IT standards in the news*. 2001. [Consulta: 6/1/2002]. <<http://www.iptc.org/site/history.html>>

HALL, K. M. *IPTC - NAA Information Interchange Model Guideline 2*. 1994. [Consulta: 29/11/2003]. <<http://www.iptc.org/download/download.php?fn=iimgl2.pdf>>

HERNÁNDEZ PÉREZ, Antonio y CARIDAD SEBASTIÁN, Mercedes. Documentación escrita en los centros de documentación de los medios de comunicación. En: MOREIRO, José Antonio (Coordinador.). *Manual de documentación informativa*. Madrid: Cátedra, 2000, p. 37-82.

HERSH, W. R. y HICHAM, D. Information retrieval in medicine: The SAPHIRE experience. En: *Journal of the American Society for Information Science*, vol. 46, nº 7, p. 743-747.

IDEALLIANCE. *PRISM: Publishing Requirements for Industry Standard Metadata*. 2003. [Consulta: 29/10/2003]. <<http://www.prismstandard.org/>>

INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL. *NewsML - Markup for the Third millenium*. 2003. [Consulta: 30/6/2003]. <<http://www.iptc.org/site/NewsML/index.html>>

INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL y NEWSPAPER ASSOCIATION OF AMERICA. *IPTC - NAA Information Interchange Model Guideline 1*. 1993. [Consulta: 20/3/2003]. <<http://www.iptc.org/download/iimv4.zip>, fichero: iimgl1.pdf>

----- *IPTC - NAA Digital Newsphoto Parameter Record Guideline 4*. 1997a. [Consulta: 20/3/2003a]. <<http://www.iptc.org/download/dnprv4.zip>>

----- *IPTC - NAA Information Interchange Model Guideline 4*. 1997b. [Consulta: 12/2/2003b]. <<http://www.iptc.org/download/download.php?fn=IIMV4.1.pdf>>

INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC). *NITF - News Industry Text Format*. 2001. [Consulta: 28/1/2002]. <<http://www.nitf.org>>

----- *SportsML*. 2003a. [Consulta: 30/4/2003a]. <<http://www.sportsml.com>>

----- *The Information Interchange Model*. [Consulta: 30/4/2003b].
<<http://www.iptc.org/download/iimv4.zip>>

----- *IPTC-TopicSet-List (en castellano)*. 2003c. [Consulta: 13/12/2003c].
<http://www.iptc.org/metadata/mtdta_ts-table01.php>

----- *Metadata: Subject Reference System & NewsML TopicSets*. [Consulta: 13/12/2003d].
<<http://www.iptc.org/metadata>>

INTERNATIONAL STANDARD ORGANIZATION. *ISO 8879:1986. Information processing - Text and office systems - Standard Generalized Markup Language (SGML)*. Ginebra: International Standard Organization, 1986.

----- *ISO/IEC 7498-1:1994. Information technology - Open Systems Interconnection - Basic Reference Model: The Basic Model*. Ginebra: International Standard Organization, 1994.

KARBEN, Alan. Responding to requeriments. En: *IPTC Spectrum*, 2002, n° 17, p. 16. [Consulta: 20/1/2003]. <<http://www.iptc.org/site/mirror/spectrum17.pdf>>

LASSILA, Ora y SWICK, Ralph R. *Resource Description Framework (RDF) Model and Syntax Specification*. 1999. [Consulta: 16/8/2003].
<<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>

MARTÍN GALÁN, Bonifacio. *Tratamiento y difusión de información jurisprudencial mediante tecnologías XML: aplicación al caso del Tribunal Constitucional*. Universidad Carlos III de Madrid, 2001.

MEGGINSON, David. *XMLNews Technical Overview*. 1999. [Consulta: 28/4/2003].
<<http://www.xmlnews.org/docs/tech-overview.html>>

PEMBERTON, Steven. *XHTML 1.0: The Extensible HyperText Markup Language. A Reformulation of HTML 4 in XML 1.0*. 2000. [Consulta: 15/12/2003]. <<http://www.w3.org/TR/xhtml1>>

PILGRIM, Mark. What is RSS? En: *XML.com*, 2002. [Consulta: 30/4/2003]. <<http://www.xml.com/lpt/a/2002/12/18/dive-into-xml.html>>

RAGGETT, Dave. A history of HTML. En: RAGGETT, Dave, LAM, Jenny, ALEXANDER, Ian, y KMIEC, Michael. *Raggett on HTML 4*. Harlow: Addison Wesley Longman, 1998.

ROCHE, Kevin. *IPTC - NAA Information Interchange Model Guideline 3*. Windsor: IPTC, 1997. [Consulta: 2/12/2003]. <<http://www.iptc.org/download/download.php?fn=IIMV3.PDF>>

SYKES, Jan. *The value of indexing*. 2001. [Consulta: 6/5/2003]. <<http://www.factiva.com/infopro/indexingwhitepaper.pdf>>

THE JAPAN NEWSPAPER PUBLISHERS & EDITORS ASSOCIATION (NSK) y International Press Telecommunications Council. *ProgramGuideML (Radio & TV Program Listing Markup)*. 2003. [Consulta: 15/12/2003]. <<http://www.programguideml.org>>

WEINBERG, Bella Hass. *Word frequency and automatic indexing*. Columbia University (New York), 1983.

----- Complexity In Indexing Systems - Abandonment And Failure: Implications For Organizing The Internet. En: *Global Complexity: Information, Chaos and Control. ASIS 1996 Annual Meeting*, Baltimore. Silver Spring: ASIS, 1996, p. [Consulta: 12/5/2003]. <<http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>>

Capítulo 5

DESARROLLOS CONCRETOS DE LENGUAJES PARA MERCADO PERIODÍSTICO: DE RSS A SPORTSML

5.1. Lenguajes basados en metadatos

La primera alternativa de uso del marcado aplicado a documentos periodísticos se limita, fundamentalmente, a la descripción de cada documento como un conjunto, sin hacer hincapié en su estructura. Esta opción ha sido fomentada con el uso de lenguajes elaborados mediante variaciones basadas en el metalenguaje RDF – fundamentado a su vez en XML–.

RDF consiste en una gramática para la elaboración de documentos secundarios que contienen metadatos fácilmente elaborables. RDF no define cuáles serán esos posibles metadatos, sino cómo puede ser la estructura física de los documentos que contengan esos metadatos, de modo que esos documentos puedan ser legibles y reutilizables por otras personas o aplicaciones informáticas.

Basándose en RDF, pueden generarse nuevos lenguajes específicos que definen el nombre de los posibles metadatos y el posible valor que pueden contener. Al estar RDF basado en XML, esos nuevos lenguajes pueden, a su vez, combinarse entre sí.

Por ejemplo, puede partirse de la versión RDF de un lenguaje de metadatos de propósito general, como el definido por la iniciativa de metadatos Dublin Core (BEGED-DOV, 2000b), que contiene elementos con metadatos aplicables a cualquier recurso en general, como el autor y el título. Y estos elementos pueden luego añadirse a los elementos de metadatos propios de un lenguaje específico. Un ejemplo aplicable a documentos periodísticos: puede definirse un lenguaje propio para documentos con metadatos periodísticos, que incluya un elemento con el metadato del título del documento, y añadirle el elemento correspondiente de Dublin Core para incluir al autor del documento. Técnicamente, esto es posible mediante la definición de DTD adicionales o de espacios de nombre (*namespaces*) en los documentos XML.

5.1.1. RSS

RSS es un lenguaje definido, desde sus orígenes, para la realización sencilla de documentos secundarios, esto es, para mostrar información sobre recursos (documentos primarios accesibles en la web) mediante documentos con metadatos físicamente separados de esos recursos.

Los documentos RSS sirven para orientar a los usuarios sobre la existencia de los recursos y para describir de modo más o menos simple su contenido su la ubicación física de los mismos. Esos documentos RSS pueden ser compartidos entre distintos sitios web, de modo que desde uno o varios de ellos se avise de la existencia de los recursos disponible en otro sitio web. Un posible uso sería la inclusión en un sitio web de un documento RSS con los titulares de una publicación accesible en otro sitio web diferente.

La elaboración de documentos RSS es bastante sencilla. Paradójicamente esa sencillez ha complicado su evolución posterior: esas siglas dan lugar hoy a dos versiones del mismo lenguaje que poseen incluso nombres distintos y formas distintas de elaboración, aunque su utilidad siga siendo la misma.

5.1.1.1. Breve sinopsis histórica: un lenguaje, dos versiones

RSS fue diseñado en 1999 por Netscape para la difusión de titulares en sitios distintos a aquellos en los que se redactaban las noticias de las que formaban parte. Al abandonar Netscape esa actividad en un corto espacio de tiempo, dejó también de fomentar el lenguaje; pese a ello, diversos distribuidores de contenidos decidieron continuar con su desarrollo.

Lo que inicialmente fue un esfuerzo conjunto se convirtió pronto en dos corrientes de opinión. La primera de ellas abogaba por mantener la simplicidad del conjunto inicial, muy reducido, y permitir la implementación de módulos que permitieran, a quien lo quisiera, hacer más complejo el uso de elementos, sin que ello impidiera

usar el esquema básico a cualquiera, con lo cual los documentos anteriores seguirían siendo legibles.

La segunda opción abogó por rechazar de plano el uso de módulos y fomentar la creación de nuevos elementos añadidos directamente a la especificación de RSS. Dado que no hubo acuerdo entre ambas tendencias, se han desarrollado desde entonces dos versiones de RSS. La primera denominada RSS 1.0 (BEGED-DOV, 2000a), data de 2000, y deriva directamente de la versión original, la 0.9, y mantenía su base en RDF: la transcripción de las siglas, es, en este caso, *Resumen de sitios mediante RDF (RDF Site Summary)*. La segunda versión evolucionó hacia una nueva versión, la 0.92, para acabar actualmente en la versión 2.0 (WINER, 2002), surgida en 2002, y cuya especificación transcribe las siglas como *Sindicación Realmente Simple (Really Simple Syndication)*. Ambas versiones coexisten sin que, desde cada una de las respectivas especificaciones se refiera a la otra.

La diferencia más notable entre ambas, además de la denominación, es que RSS 2.0, al igual que RSS 0.92, ha abandonado el uso de RDF y se mantiene como un lenguaje basado directamente en XML, además de incluir una veintena adicional de elementos opcionales. RSS 2.0, al estar basado en XML, admite expresamente el uso adicional de cualquier elemento que no sea RSS 2.0, mediante el uso de espacios de nombre, lo cual no está muy lejos de la posibilidad de implementar módulos adicionales de elementos, que motivó en parte la división del lenguaje en dos versiones diferentes.

5.1.1.2. Estructura básica y elementos opcionales

Tanto la versión 0.91 y 1.0 como la 2.0 comparten una estructura básica, dividida esencialmente en dos mitades: una primera, que especifica un sitio web, o **canal** en la terminología RSS, conteniendo elementos que le dan un título, especifican su dirección en la web y describen brevemente el mismo, además de contener opcionalmente una imagen; la segunda parte del documento contiene referencias a cada uno de los **ítems**, o documentos incluidos en ese canal de los que se contienen metadatos, indicando para cada uno de ellos otra tríada con su título, la dirección

concreta de cada documento y una breve descripción. En RSS 1.0, además, la primera parte contiene, duplicadas, las referencias a las direcciones de cada uno de los sitios.

Como hemos indicado, en RSS 1.0, estos elementos están basados en RDF, mientras que en RSS 2.0, se trata de elementos XML; sin embargo, en ambos casos, los nombres y las funciones son exactamente los mismos. Un documento RSS en ambas versiones puede contener tantas descripciones de documentos distintos como se desee.

RSS 1.0 permite además el uso de elementos opcionales de distintas categorías, incluidos los elementos de Dublin Core. Por su parte, RSS 2.0 permite, además, el empleo de elementos opcionales tanto para definir el canal como cada uno de los ítems. En el caso del canal, RSS 2.0 permite especificar el lenguaje humano de los contenidos incluidos en el sitio, menciones sobre el copyright, el responsable intelectual del sitio referido y el webmaster, la fecha de publicación inicial y de actualización del sitio, la categoría temática –ya sea en lenguaje libre o en lenguaje controlado, especificando dónde se encuentra ese lenguaje controlado– y otros metadatos técnicos que faciliten el envío de información entre el sitio web productor del recurso referido y el sitio web que contenga los metadatos creados. RSS 2.0 incluye también metadatos opcionales sobre cada ítem: la dirección de correo electrónico del autor, la categoría en lenguaje libre o controlado, la fecha de creación, la dirección de documentos opcionales con comentarios, recursos multimedia u otros sitios RSS relacionados, etc.

5.1.1.3. Rasgos documentales de RSS

RSS constituye un lenguaje exclusivamente de descripción de contenidos, que permite mostrar en un solo documento información breve sobre uno o varios recursos: aplicado a documentos periodísticos, información sobre los titulares de una o varias noticias, mención de sus autores, de la dirección de la página web en la que se encuentra cada noticia, un breve resumen de la misma y, en el caso de la versión 2.0, algunos datos adicionales, como el lenguaje de la noticia, la categoría temática

(en texto libre), etc. El objetivo de RSS, como ya se indicó en el capítulo 2, consiste en permitir que las menciones a las noticias de un medio sean visibles desde un sitio web ajeno a ese medio.

Su capacidad documental se reduce a la de ser un primer punto de acceso a una serie de noticias, basándose en unos breves descriptores básicos. RSS, pues, es un formato de salida para fichas descriptivas sobre noticias, pero no es un formato pensado para almacenar información periodística, a medio y largo plazo. RSS es más un lenguaje pensado exclusivamente para la sindicación gratuita de contenidos, de un modo sencillo. Como indica uno de sus desarrolladores, Raël Dornfest, citado por (BANNAN, 2002),

“it's very easy. The content provider owns the hits and the content. It's like free advertising”.

Los documentos RSS suelen tener una duración muy concreta, pretendiendo mostrar las últimas informaciones que ha realizado el medio; más allá de una duración determinada, los documentos RSS no son almacenados, aunque sí es posible emplear herramientas informáticas que, siguiendo las palabras de Winer también citadas por (BANNAN, 2002),

“there are centralized applications and server applications that read the RSS source [de los ficheros ofrecidos por diversos medios] and generate information on topics that you're interested in”.

Sin embargo, las propias carencias de RSS como herramienta para uso documental son manifestadas explícitamente por el citado Dornfest en el mismo artículo:

“For me, an RSS subject could be a path or a URL. For someone else, it's keywords. Certain things mean different things for different people”.

El hecho de que RSS 1.0, al estar basado en RDF, permita el uso de *módulos*, es decir, de posibles lenguajes de metadatos que permitan una mayor precisión y normalización documentales, como el basado en Dublin Core, o de que RSS 2.0,

además de usar también módulos mediante espacios de nombre, contenga un mayor número de etiquetas, puede aumentar la posibilidad de un uso documental de RSS.

No obstante, esta posibilidad choca con dos condicionantes. El primero radica en que también serían formalmente válidos (aunque mucho menos operativos documentalmente) documentos RSS que se ajustaran a las especificaciones mínimas: un título, una dirección de una página y una descripción; tampoco ninguno de estos elementos resulta siquiera obligatorio.

De hecho, RSS es más útil como un formato de sindicación que como un formato meramente documental: sí puede ofrecer utilidad, por ejemplo, como un formato de exportación de titulares, como se comentará en el capítulo 4, a partir de noticias generadas en formatos más amplios, de los que seleccionar información básica y convertirla a RSS.

El segundo condicionante afecta a la capacidad de que esos módulos se consoliden como realidades utilizables. Buena parte de los mismos son únicamente propuestas (SWARTZ, 2003), con lo cual resulta difícil que la industria periodística, e incluso que los pequeños productores de información periodística, apuesten por lenguajes que no tengan un mínimo de estabilidad ni de asunción por parte de un colectivo. De ahí que los usuarios de RSS, en la actualidad, ya sea de su versión 1.0, como (EL MUNDO.ES, 2003), o de la versión 2.0, como (NEWS.COM, 2003), opten por emplear las opciones más básicas de la respectivas versiones.

De hecho, en la actualidad, se está promoviendo una remodelación de este lenguaje, denominada provisionalmente como Atom (NOTTINGHAM, 2003).

5.1.2. PRISM

El otro lenguaje de marcado dedicado, que permite la descripción de un contenido físicamente separado mediante metadatos es PRISM, abreviatura traducible por Metadatos Requeridos para la Industria Editorial (Publishing Requirements for Industry Standard Metadata). PRISM no es en realidad un único lenguaje, sino un

conjunto de lenguajes de metadatos destinados al intercambio de productos realizados por todo tipo de empresas editoras de contenidos, entre las cuales, pero no exclusivamente, se encuentran editoras de prensa de ámbito general, especialmente de revistas, o especializadas en la elaboración y difusión de publicaciones científicas, tanto de revistas como de libros, así como distribuidores de bases de datos de prensa, otras empresas editoras de diferentes contenidos y empresas dedicadas a la gestión del conocimiento. Todas ellas están agrupadas en el consorcio IdeAlliance, que se ha dedicado en los últimos años a la elaboración de diversos lenguajes basados en XML.

Las dos grandes utilidades de PRISM son la descripción de recursos conjuntos (esto es, de documentos concretos) con propósito general, así como el establecimiento de relaciones entre estos recursos, con el fin de facilitar su reutilización.

De entre el conjunto de lenguajes de PRISM, el principal, denominado con el mismo nombre, puede contener a los otros mediante el uso de espacios de nombre; de este mismo modo, puede contener además elementos de cualquier otro lenguaje basado en RDF. De hecho, el lenguaje principal PRISM parte de la definición de elementos en RDF elaboradas por Dublin Core, amplificadas con elementos propios, como se describe más adelante.

PRISM incluye otros lenguajes propios: uno para definir el uso de vocabularios controlados, denominado PCV (Vocabulario Controlado de PRISM); un lenguaje básico para definir datos que describen derechos de uso de un documento, denominado PRL (Lenguaje de Derechos de PRISM) y, finalmente, un lenguaje destinado al marcado de contenido inserto en el documento, denominado PIM (Marcado en línea de PRISM). Como veremos en un apartado posterior, estos dos últimos son realmente sendas propuestas más que lenguajes plenamente desarrollados.

Asimismo, PRISM ha definido una serie de listas de autoridades, cuyos valores son usados normativamente como contenido de los distintos elementos de metadatos de PRISM, que forman parte asimismo de sendos lenguajes también definidos por esta

iniciativa. Estos lenguajes, destinados a expresar géneros periodísticos o editoriales, tipos físicos de recurso, derechos de uso, posibilidad de utilización de un documento y, por último, derechos de uso, serán referidos al hacer mención de los elementos en los que puedan insertarse.

5.1.2.1. El lenguaje básico PRISM

En síntesis, un documento PRISM puede contener tanto los elementos de Dublin Core como otros definidos por la citada especificación PRISM (DANIEL; HANSEN, y POPE, 2003b), que pueden agruparse en varios bloques:

- Por una parte, los **elementos de propósito general**, que contienen datos básicos para identificar al documento referido. Todos ellos (código identificador del documento, título, autor, lenguaje, etc.) son elementos Dublin Core, excepto dos propios de PRISM. El primero describe la *categoría editorial*, o género en el que se engloba el documento. El valor de este último metadato debe ser uno de los términos de una lista de autoridades con categorías o géneros definidas expresamente en la especificación de PRISM. En el caso de los documentos periodísticos, se puede optar entre categorías como breve, columna, entrevista, noticia, obituario, etc.
- El segundo elemento propio de PRISM en este apartado describe el *tipo de recurso*, entendido como el “estilo” de presentación física: artículo, libro, revista, periódico, página web... pero también lista, tabla, índice, barra... e incluso, en el caso de que sea una imagen, posibilidades como “ojo de pez”.
- Tanto las categorías editoriales como los tipos de recursos, incluyendo sus definiciones concretas, están definidas en parte por la propia especificación PRISM o han sido expresamente tomadas de diversos documentos, incluido el Tesoro Getty de Arte y Literatura (GETTY RESEARCH INSTITUTE, 2003) o, en algunos casos, de las especificaciones de NewsML.

- Otro conjunto de elementos está formado por aquellos que designan aspectos relacionados con el **origen y el soporte del documento descrito**. Dado que, en buena parte, los documentos originales para los que se pretende destinar la elaboración de metadatos PRISM son documentos impresos, se incluyen detalles como el número, el volumen, el ámbito de la edición o la página inicial, junto a la sección del medio en la que se encuentra un documento, cuando se trata de un texto incluido en una publicación periódica.
- Un tercer conjunto de elementos son empleados para la **descripción temática**, como la mención a una materia o a la cobertura temporal del documento (por ejemplo, “el siglo XX”) mediante sendos elementos Dublin Core, a los que se añaden elementos de metadatos propios de PRISM para indicar un hecho concreto, un sector industrial, una localización geográfica, una persona concreta, una organización de cualquier tipo o una sección dentro de un medio informativo en la que se ha incluido ese documento.
- Especialmente importante en PRISM es el conjunto de elementos que pretenden explicar las posibles **relaciones de un documento** con otros físicamente independientes. PRISM permite especificar que un documento es parte de otro, o viceversa; que un documento dispone de versiones, intelectuales o de formato, en otros documentos; que el documento realiza referencias a otros documentos o es referido por éstos; que se deriva, es una adaptación, una traducción o una corrección de otro documento, o viceversa; que el documento requiere, físicamente, de otro fichero, o viceversa; y finalmente, que el documento tiene una versión alternativa en otro documento, o al revés. Como puede verse, algunas de las distinciones realizadas, cada una de ellas mediante sendos elementos, presentan una notable sutileza semántica.
- Finalmente, el lenguaje básico de PRISM ofrece elementos a modo de metadatos con las **fechas** de creación, modificación, distribución a otro

medio, publicación en otro soporte y fin de la validez de un documento (por ejemplo, en su cesión a otro medio).

5.1.2.2. Lenguajes documentales en PRISM

La propia especificación de PRISM define, en su parte no normativa, dos tipos de **lenguajes controlados** que pueden ser usados en los diferentes lenguajes de PRISM: una lista de autoridades, cuyo ejemplo puede ser la citada lista de géneros documentales aplicada a los documentos PRISM; y por otra parte, lo que en la especificación de PRISM se denomina *taxonomías*, que se definen literalmente como

“una de las formas de vocabulario controlado, en la cual los únicos conceptos identificados están ordenados en una jerarquía que representa las relaciones entre conceptos más generales y más específicos.”
(DANIEL; HANSEN, y POPE, 2003a)

Esta definición, y el Lenguaje de Vocabularios Controlados PRISM (PCV), que define los elementos empleados para estructurar estos vocabularios, remite a lo que en la terminología documental tradicional se entiende como una *lista de encabezamiento de materias*. La lista será un documento externo, al que se podrá acceder mediante un URN o un URL que contendrán la dirección del término exacto dentro de la lista. PCV permitirá además que, junto al URI concreto, se incluyan como elementos adicionales de metadatos un código de identificación único, la mención del vocabulario y del término usados, mediante texto legible por personas, una definición del término (equivalente a una nota en las listas de encabezamiento) y sendos elementos para mencionar términos genéricos y específicos del término empleado, así como sinónimos y términos relacionados.

Asimismo, PRISM ofrece un breve lenguaje de marcado para la descripción de ciertos derechos de uso sobre un documento. Este lenguaje, denominado PRL, también es utilizable en forma de metadatos externos, y añade, a los metadatos concretos de PRISM sobre derechos (el elemento `dc:rights`, proveniente de Dublin Core, así como elementos sobre el copyright y la mención de la agencia que tiene encomendada la gestión de esos derechos), tres elementos de un pequeño

Lenguaje de Derechos de PRISM (PRL), que restringen la aplicación de esos derechos a un área geográfica concreta, limitan los temas junto a los que puede relacionarse ese documento (por ejemplo, si se trata de un artículo, su uso puede estar supeditado a que no aparezca en una sección concreta o junto a una información de un tema determinado) y, finalmente, indican la posibilidad de uso de ese documento, mediante valores de una lista de autoridades, también definida por PRISM, indicando si el documento es utilizable, no lo es, o no se tiene constancia de permiso o prohibición alguna.

5.1.2.3. Otros lenguajes de marcado PRISM

La iniciativa PRISM ofrece otros dos lenguajes, en este caso, para permitir el empleo de marcado inserto

El primero de ellos es el Lenguaje de Marcado en Línea de PRISM (PIM), consistente en un breve listado de elementos PRISM que pueden ser empleados como elementos de marcado inserto dentro de un texto (a diferencia del resto de los lenguajes de la familia PRISM ya mencionados, cuyos elementos están pensados inicialmente para ser incluidos en un documento externo o bien, de modo adjunto al documento que describen). Los elementos PIM son utilizados para destacar referencias a personas, hechos, sectores industriales, localizaciones geográficas, organizaciones de cualquier tipo, títulos de obras intelectuales o, finalmente, citas concretas. Salvo los dos últimos, los anteriores son equivalentes a otros elementos similares, descritos como metadatos temáticos del lenguaje general de PRISM aplicados a la descripción temática.

El segundo de los lenguajes de marcado inserto PRISM, el más reciente en su creación, es el PRISM Aggregator Message (PAM), o lenguaje que permite el intercambio expreso de documentos entre los productores de los mismos y los agregadores que distribuyan el texto completo de los contenidos. PAM, como veremos más adelante, presenta por ello ciertas similitudes con NITF.

Fundamentalmente, un documento PAM, definido por una serie de DTD explicadas mediante la correspondiente especificación ((IDEALLIANCE, 2003a)), consta de un elemento principal o raíz llamado `pam:message`, que a su vez puede contener uno o varios elementos `pam:article`; cada uno de estos últimos corresponde a un documento periodístico original; es decir, un documento PAM puede contener un paquete de piezas documentales distintas –por ejemplo, varios artículos publicados en una revista–.

Un `pam:article` contiene dos elementos: `head` y `body`. Como puede observarse, estos elementos están basados en los de igual nombre en HTML: de hecho, su uso es similar. El elemento `head` contiene los metadatos, basados en los distintos lenguajes PRISM de marcado adjunto (los citados PRISM, PRL, PCV, etc.) que se pretendan asignar a la pieza correspondiente.

El elemento `body`, por su parte, incluye el contenido de la pieza, incluido mediante elementos basados en un subconjunto de XHTML, del cual se han eliminado expresamente algunos elementos como los marcos o frames.

Algunos de estos elementos pueden contener atributos que sirvan, por ejemplo, para indicar una utilidad concreta de ese elemento. Por ejemplo, si se quiere indicar que un párrafo del contenido incluye una referencia al autor del mismo, se emplearía el elemento `p` con un atributo `class` cuyo valor, obligatoriamente, sea `byline`.

PAM puede también incluir elementos de marcado de PIM. Un ejemplo de uso de este lenguaje puede verse en el documento de la siguiente figura:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE pam:message PUBLIC
"-//PRISMstandard.org//DTD Aggregation with XHTML v1.0//EN">
<pam:message xmlns:pam=
"http://prismstandard.org/namespaces/pam/0.0/"
xmlns:prism="http://prismstandard.org/namespaces/basic/1.2/"
xmlns:pim="http://prismstandard.org/namespaces/pim/1.2/"
xmlns:dc="http://purl.org/dc/elements/1.1/">

<pam:article>
<head>
<dc:identifier>335398</dc:identifier>
<dc:title>The Gong Goodbye</dc:title>
<dc:creator>Karen Valby</dc:creator>
```

```
<prism:publicationName>ENTERTAINMENT WEEKLY
</prism:publicationName >
<prism:coverDate>2003-01-10</prism:coverDate>
</head>
<body>
<h1>The Gong Goodbye</h1>
<p class="deck">Was game-show king <pim:person>Chuck Barris
</pim:person> a hitman or just a hitmaker? The subject of
<pim:objectTitle>Confessions of a Dangerous Mind
</pim:objectTitle> confronts his past lives.</p>

<p class="byline">Karen Valby</p>
<p>Chuck Barris can't stand the sight of himself. Old
  episodes of The Gong Show, the daffy '70s talent show
  he created and hosted with manic glee...."</p>
</body>
</pam:article>
</pam:message>
```

Figura 5.1.
Esquema de un posible documento PAM.

Adaptación basada en (IDEALLIANCE, 2003a)

El ejemplo muestra un documento PAM con una sola pieza, pero puede contener tantas como se desee. En esa pieza, se observa que los párrafos, definidos con el elemento `p`, incluyen varios tipos de atributos. Uno de ellos, `deck`, indica que ese párrafo funciona como una suerte de entradilla; otro atributo, el citado `byline`, indica que en el texto se incluye la mención del autor del documento.

En el ejemplo, también se muestra como algunas referencias detalladas dentro del texto han sido destacadas mediante marcado inserto, utilizando elementos del lenguaje PIM. Así ocurre con `pim:person`, que indica que se hace una mención a una persona concreta, o con `pim:objectTitle`, que señala una mención a una obra concreta (literaria, en este caso).

5.1.2.4. Rasgos documentales de PRISM

PRISM, al contrario que RSS, es un lenguaje mucho más complejo y, de hecho, su capacidad de aplicación documental resulta evidente. Como ya hemos indicado, se trata de una iniciativa que combina varios lenguajes de metadatos, así como la posibilidad, de emplear marcado inserto, adjunto o externo.

En la definición de PRISM participan expresamente, entre otras entidades, empresas dedicadas a la elaboración y aplicación de lenguajes de clasificación (que denominan *taxonomías*), lo cual se traduce en la creación de diversos mecanismos para describir todo tipo de objetos, ya sea desde un punto de vista técnico, o bien, refiriéndose a su contenido. De hecho, todas las explicaciones sobre la iniciativa (como, por ejemplo, (BURMAN, 2002)) tienen un fuerte, y casi específico, trasfondo documental.

Desde el punto de vista documental, PRISM cuenta con dos herramientas destacables. Una de ellas es el conjunto de vocabularios definidos para describir los tipos de objetos susceptibles de ser descritos mediante metadatos, tanto desde el punto de vista físico (por ejemplo, variados tipos de fotografía, texto impreso, página web, etc.) como desde el punto de vista del género o clase intelectual empleado (en lo que atañe al periodismo: artículo, crónica, reportaje, entrevista, etc.).

La segunda gran herramienta documental de PRISM que debe destacarse es el subconjunto de su lenguaje principal dedicado a la definición de relaciones entre diferentes documentos, que permite indicar, por ejemplo, si un documento es parte de otro o si está dividido en varias partes, y dónde está cada una; si el documento tiene referencias de otros, o si es referido por otros documentos; si el documento tiene versiones físicas o intelectuales, de varios tipos, etc.

No obstante, PRISM presenta también cuatro detalles que deben ser tomados en cuenta a la hora de considerar su posible uso documental con fines periodísticos.

El primero de ellos es que, al igual que RSS, PRISM describe fundamentalmente documentos completos, sin la posibilidad de tener en cuenta sus piezas u objetos por separado, salvo que esos objetos estén físicamente en ficheros separados (en cuyo caso, sí es posible relacionar esos objetos mediante los citados elementos de relación). PRISM no dispone de la capacidad de describir la estructura de los mismos ni de usar, esa información desde un punto de vista documental.

Por otra parte, el citado PIM (Marcado en línea de PRISM), apenas está formado por media docena de elementos, señala de modo muy básico algunos detalles dentro de un texto (la mención a personas, empresas, etc.), es aún solo un borrador, y tampoco

se ha especificado aún cómo será su uso entre las diversas posibilidades existentes: en un lenguaje propio, como parte de otro vocabulario mayor, etc.

El segundo factor se refiere a los vocabularios controlados antes citados sobre formato físico e intelectual, cuentan con el respaldo de instituciones con experiencia en la descripción documental; un ejemplo sería el citado *Tesaurus Getty de Arte y Arquitectura* ((GETTY RESEARCH INSTITUTE, 2003)). Sin embargo, tales términos, debido a la propia naturaleza de PRISM –un conjunto de lenguajes para la descripción general de publicaciones de todo tipo–, no son necesariamente de uso periodístico.

El tercer factor reside en el hecho de que la iniciativa PRISM se base fundamentalmente en desarrollos externos ya definidos y empleados por separado, como RDF, Dublin Core, RSS o XHTML, puede facilitar su utilización, dado que esos lenguajes ya son previamente conocidos.

A cambio, este uso asume también algunas de las debilidades de todos ellos: la excesiva generalidad de su descripción para su aplicación a un campo concreto, en el caso de Dublin Core ((DUBLIN CORE METADATA INITIATIVE, 2002)), aunque trate de paliarla con la definición de metadatos específicos aplicables a contenidos publicados por una empresa editorial, periodística o no–; la excesiva simplicidad y la falta de un desarrollo adecuado para la definición de lenguajes controlados o para el almacenamiento de contenidos, en el caso de RSS; y la floja definición estructural que es posible con XHTML, aunque también se pretenda paliar mediante el uso de atributos específicos que distingan diferentes usos de un mismo elemento, como se acaba de mostrar con el uso del elemento `p` en la figura 5.1.

Por otra parte, los desarrollos propios de PRISM, que amplían las carencias de cada uno de los lenguajes en los que se basan, son en muchos casos muy semejantes, e incluso de menor profundidad documental en la mayoría de las ocasiones, que los empleados en otros lenguajes con los que comparte pretensiones de descripción documental, como NITF o NewsML.

PRISM es asimismo un lenguaje expresamente interrelacionado con otros lenguajes de marcado dedicados a la descripción documental: aparte de reconocer los evidentes lazos con NITF y NewsML (DANIEL; HANSEN, y POPE, 2003b, p. 10) –que han desembocado, de palabra, en la incorporación de miembros del IPTC en el comité de desarrollo de este lenguaje–, PRISM está basado originalmente en el uso de RDF, lo cual incluye a esta iniciativa en el ámbito de la web semántica, hasta el punto de ser citada expresamente por el propio W3C. Uno de sus posibles usos, el desarrollo hacia la implementación de PRISM como un módulo adicional dentro del lenguaje RSS basada en la versión 1.0 de este último, parece apuntar en esta idea. (HAMMOND; HANNAY, y LUND, 2003).

Todas estas conexiones suponen una incertidumbre sobre los motivos que pueden decidir al uso de PRISM, de los otros posibles lenguajes o de una combinación de todos ellos. De hecho, PAM, como hemos mencionado, presenta ciertas coincidencias en su posible uso con NITF y, asimismo, con NewsML, lo cual induce a la búsqueda de relaciones más profundas entre esta iniciativa y esos otros lenguajes.

Un cuarto factor que puede arrojar una luz sobre el futuro del uso de PRISM puede ser la aparición de una DTD concreta que permite poner en práctica los diferentes lenguajes de la iniciativa, mediante el uso del citado PAM. A favor de éste último juegan también los anuncios de la propia (IDEALLIANCE, 2003b), según la cual varios de los medios que apoyan esta iniciativa han decidido utilizar los lenguajes ya definidos de un modo concreto: actores del tamaño de la editora Time Inc. (una de las mayores empresas de difusión periodística del mundo) y los agregadores Lexis-Nexis y ProQuest (que estudian el uso conjunto de PAM para el intercambio de noticias a partir de "finales de 2003"), así como la editora norteamericana Hearst Publishing, que posee una cincuentena de medios periodísticos de todo tipo, y la empresa Platts, subsidiaria de McGrawHill dedicada a la difusión de información sobre energía que se autodenomina "el mayor proveedor mundial de información sobre energía".

El tamaño de estos medios puede suponer una ventaja, dada su fuerza, pero también puede suponer un inconveniente, si otros medios contemplan esta y otras iniciativas como una forma de imponer el modelo de trabajo de estos grandes conglomerados.

5.2. Lenguajes de marcado semántico-estructural

Frente a la opción descriptiva mediante metadatos, una segunda alternativa es la formada por la familia de lenguajes que permiten el doble marcado, semántico y estructural, de informaciones periodísticas. Si PRISM, mediante la variedad PAM, ya permite esta opción, no desarrolla todas sus posibilidades.

De hecho, en buena medida el modelo de PAM está basado en la familia de lenguajes volcados sobre la definición y descripción estructural de los documentos periodísticos, sin olvidar la descripción documental aplicada a los mismos. Esta familia, definida por el organismo internacional IPTC, está compuesta fundamentalmente por tres lenguajes. Mencionados por orden de especificidad inverso, es decir, de más genérico a más concreto, esos lenguajes se denominan NewsML, NITF y SportsML.

Aunque cada uno de ellos cumple cometidos distintos, todos ellos se basan en una serie de peculiaridades comunes que permiten relacionarlos, no solo desde un punto de vista conceptual, sino también físicamente:

- a) **Capacidad de estructuración.** Los tres lenguajes están basados en la definición de la estructura que pueden tener diversos documentos periodísticos. Esta utilidad estructural no va en detrimento de las posibilidades de descripción semántica sino que, al contrario, en muchas ocasiones, ambas se potencian entre sí.
- b) **Modularidad.** Enlazando con la característica anterior, los documentos en estos tres lenguajes están formados por una serie de módulos: unos pocos

de ellos, básicos, a los que pueden añadirse otras partes complementarias, muchas de ellas, intercambiables entre sí.

Pero la modularidad relaciona además los tres lenguajes como un conjunto o *familia*, al estar diseñados para poder trabajar conjuntamente. Cada uno de ellos puede ser útil por separado, pero también pueden ser empleados dos a dos, o los tres a la vez. Dicho de otra forma: un documento físico en NewsML puede contener dentro de él un documento físico marcado mediante NITF o mediante SportsML; un documento SportsML puede contener un documento NITF y, por tanto, puede darse el triple caso de que un documento NewsML pueda contener un documento SportsML que contenga a su vez un documento NITF; o bien, que el documento NewsML contenga una pieza con un objeto NITF, y otra pieza con un objeto SportsML...

No obstante, el uso de uno de estos lenguajes no implica necesariamente el uso de los otros. Por ejemplo, NewsML, como se explica en un apartado posterior, puede contener objetos en cualquier formato y lenguaje informático.

- c) **Capacidad de descripción.** A pesar de no ser lenguajes con fines exclusivamente descriptivos, todos ellos sí tienen un fuerte grado de capacidad descriptiva semántica tanto del propio contenido de un documento como de su entorno. Esta descripción se muestra, explícitamente, en la definición de los elementos y atributos que forman cada lenguaje, muchos de los cuales, en solitario o unidos a otros, están diseñados para realizar una explicación muy detallada, incluso contextual, de los contenidos marcados. Podemos encontrar un ejemplo en el elemento `person`, que forma parte del lenguaje NITF, y que permite no solo identificar un nombre de persona, distinguiéndolo de ese mismo nombre aplicado a un ente inanimado (por ejemplo, Juan Sebastián Elcano, como nombre propio, respecto al nombre completo usado para nombrar a un barco de la Armada española, el “Juan Sebastián Elcano”), sino que además permite especificar cuál de las partes del

nombre es el nombre de pila, y cuál o cuáles son los apellidos, mediante sendos elementos adicionales.

Pero además, esa capacidad de descripción de contenido, cualidades y entorno, es fomentada explícitamente por el propio IPTC, mediante una política de generación de lenguajes de clasificación con la que intenta normalizar las descripciones básicas de las diversas noticias, que pueden aplicarse a múltiples elementos en cada lenguaje. Los ya citados códigos temáticos son el punto básico de esa iniciativa, aunque también se han elaborado lenguajes de clasificación destinados al uso específico de los medios, detallando el género o, al menos, el tipo de documento periodístico de que se trata cada pieza informativa, según una nomenclatura clásica periodística (noticia, reportaje, análisis, etc.); y además, tratando de delimitar términos referentes al tratamiento técnico y periodístico de cada pieza (por ejemplo, que se trata de noticias elaboradas o de datos en bruto, que son publicables o que aún no lo son, que tienen derechos de reproducción cedidos solo para un rango temporal o espacial concretos, etc.)

Dado el carácter modular de los lenguajes de esta familia, realizaremos su descripción siguiendo el citado orden de especificidad inverso, describiendo aquellos lenguajes que pueden ir conteniendo a su vez a otros. Comenzaremos con el más genérico, NewsML; continuaremos con NITF, un lenguaje específico en el formato, textual, pero de propósito general en cuanto a los contenidos que puede incluir; concluiremos esta parte con la descripción de SportsML, un lenguaje específico tanto en el tema para el que ha sido diseñado, la información deportiva, como en el tipo de información al cual se aplica, la descripción de datos estadísticos sobre noticias deportivas, a la que pueden añadirse contenidos periodísticos elaborados, mediante el uso recomendable de NITF.

5.3. NewsML: mercado de contenedores periodísticos

NewsML se define como un lenguaje destinado al transporte y descripción de objetos periodísticos, esto es, de soportes físicos, independientemente de cual fuera el formato de los mismos. En otras palabras: un mismo documento NewsML puede contener, simultáneamente, varias noticias en diferentes formatos textuales, cada una de ellas generada con una aplicación informática diferente, así como un conjunto de imágenes fijas o en movimiento, y diversos ficheros de sonido, también sin importar el formato.

¿Permite NewsML generar todos estos objetos diferentes? La respuesta es no. Lo que NewsML sí permite es almacenar todos esos objetos, cada uno en su formato correspondiente, mediante un lenguaje de marcado común.

Las utilidades de este enfoque (un único lenguaje de empaquetado, diferentes lenguajes de producción) son múltiples. La primera de ellas es su facilidad para el transporte: con un solo envío, un medio productor de noticias, como una agencia, puede hacer accesibles múltiples contenidos, de una sola vez, a un medio receptor. NewsML permite además incluir los contenidos completos, o bien, solo una serie de datos sobre los mismos, de modo que el receptor decida cuáles de esos contenidos son susceptibles de interés, y sepa cómo y dónde acceder físicamente a ellos.

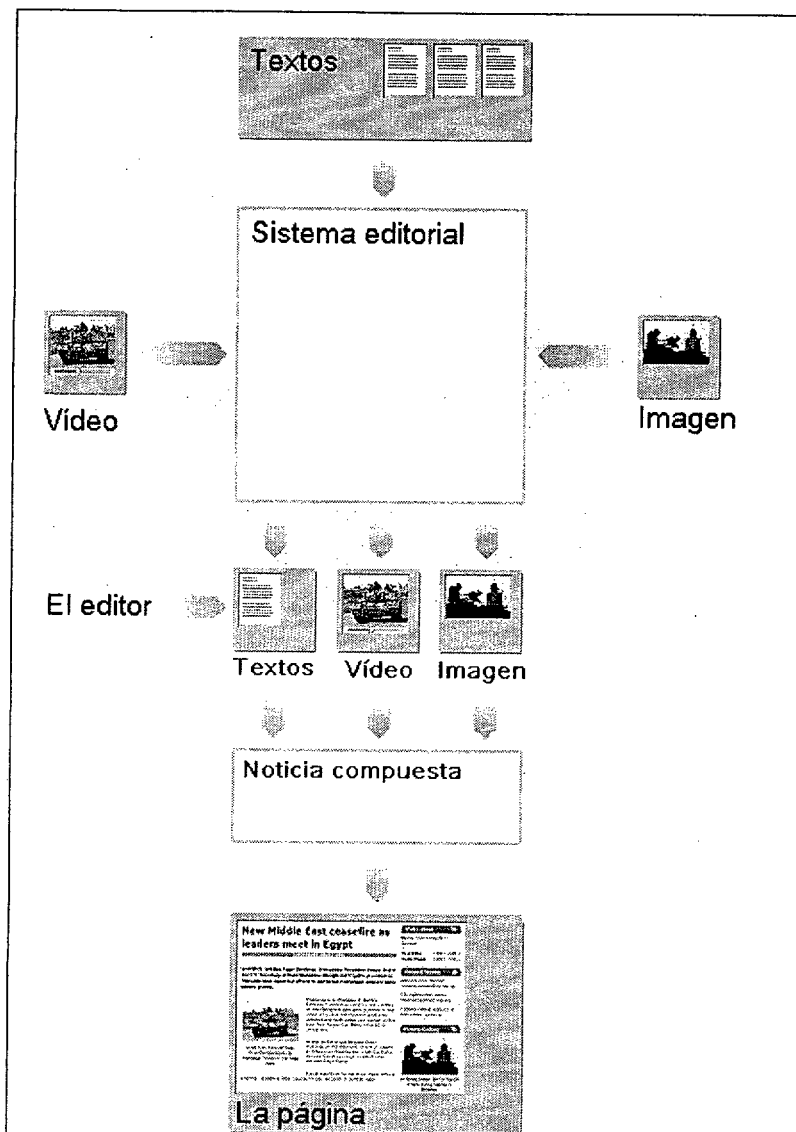


Figura 5.2.

Esquema de composición de una noticia en un sistema de edición de contenido usando NewsML

Adaptación basada en (REUTERS, 2002)

Además, NewsML permite, como se explica más adelante, definir la estructura informativa de dichos objetos: siguiendo con el caso de una noticia compuesta, puede delimitar como diferentes objetos, cada una de ellos elaborado en un formato diferente (textos, fotos, multimedia), forman parte de una misma pieza periodística, identificándolos además por separado de forma que el medio receptor pueda luego ofrecer todos esos objetos o seleccionar solo aquellos que requiera en función de su interés informativo o de condicionantes, ya sean técnicos –si se trata de una emisora

de radio, solo los sonidos; si es un periódico impreso, texto, fotografías y gráficos fijos; si se trata de un medio digital, cualquier tipo de formato y objeto— o humanos — por ejemplo, si se trata de un medio de difusión en un país concreto, solo los objetos en una o varias lenguas específicas—. Un documento NewsML puede, además, contener más de un documento periodístico a la vez.

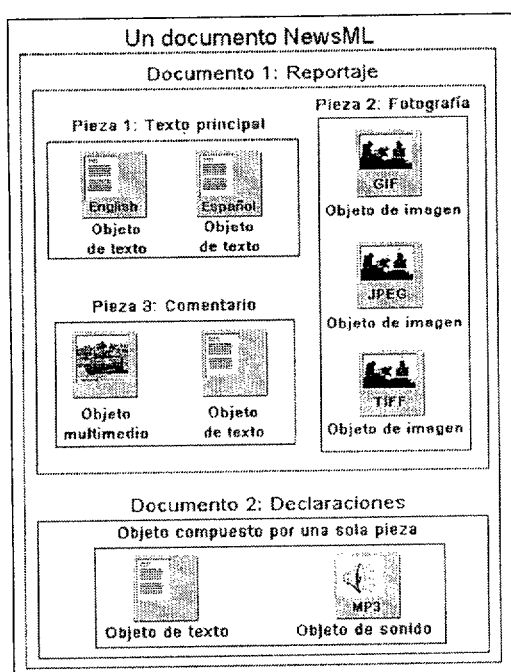


Figura 5.3.

Ejemplo de contenidos de un documento NewsML que incluye dos documentos periodísticos, cada uno compuesto por diferentes piezas y objetos

La tercera utilidad de NewsML, que determina en particular su capacidad de aplicación documental —aunque la descripción estructural también es aplicable a este uso—, es la posibilidad de añadir metadatos en cada una de las partes que forman parte de un documento NewsML: ya sea:

- Un objeto, la unidad mínima de contenido que puede incluirse en un documento NewsML.
- Un objeto repetido físicamente en diferentes formatos.
- Un contenedor que incluye al menos una información descrita mediante diferentes objetos (véase la figura 5.2.).

- Un conjunto que contiene varios contenedores, cada uno de ellos con información sobre un hecho distinto.

Esta múltiple capacidad de descripción se aplica a casi todas las posibilidades documentales: además de la definición estructural, la mención de todo tipo de cualidades periodísticas, técnicas y documentales, así como la explicación contextual de cada objeto en el conjunto del documento. Pero NewsML no permite describir, por ejemplo, un texto completo hasta sus mínimas partes, sino únicamente indicar que incluye un objeto, indicar qué tipo de objeto (textual, imagen, etc.), y aplicarle como tal las descripciones posibles en su conjunto.



Figura 5.4.
Ejemplo de la posible ubicación de metadatos en un documento NewsML

La descripción interna y detallada de cada una de las partes del objeto queda, si es posible, para el lenguaje en el que se defina dicho objeto. Por este motivo, IPTC recomienda que los contenidos textuales que formen parte de un documento NewsML sean realizados mediante NITF. No obstante, este deseo es únicamente una recomendación, dado que el funcionamiento de NewsML, es independiente del formato en el que estén elaborados los contenidos que empaqueta.

También es posible, aunque de un modo limitado, establecer relaciones entre objetos dentro de un documento NewsML, o bien, entre distintos documentos NewsML. Esta capacidad de relación facilita a los sistemas de gestión de contenidos la elaboración de instrumentos que mejoran la contextualización de uno o varios documentos, asociándolos a otros, por ejemplo, con distinta fecha de elaboración pero con contenidos comunes.

5.3.1. Estructura básica de los documentos NewsML

La estructura básica de NewsML no ha sufrido excesivas modificaciones a partir de la que, tras la propuesta inicial de la agencia Reuters, fue aprobada por IPTC en octubre de 2000. Actualmente, la versión de NewsML en vigor es la 1.1 (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2002b), en formato DTD, y su equivalente, la 1.1.2 (IPTC lanza NewsML, 2000), en formato esquema.

Un documento NewsML puede contener uno o varios elementos `ContentItem`, cada uno de los cuales incluye el contenido de un objeto periodístico (un texto, una imagen, una grabación sonora, etc.). Uno o varios `ContentItem` pueden formar parte de un elemento `NewsComponent`, que vendría a ser un acontecimiento concreto. Por ejemplo, si una agencia quisiera transmitir dos versiones en dos lenguas distintas del mismo hecho, cada una de las versiones estaría en un `ContentItem`, mientras que las dos versiones estarían dentro de un `NewsComponent`.

Si ese `NewsComponent` fuera, por ejemplo, la crónica sobre un acontecimiento y formara parte de una información compuesta físicamente por más de una pieza, cada una de las piezas iría a su vez dentro de otros `NewsComponent`. Y el conjunto de `NewsComponent` formaría parte de un `NewsItem`, de una información sobre un tema. Un documento NewsML puede contener más de una noticia, es decir, más de

un `NewsItem`. La siguiente figura muestra un posible ejemplo de un documento NewsML.

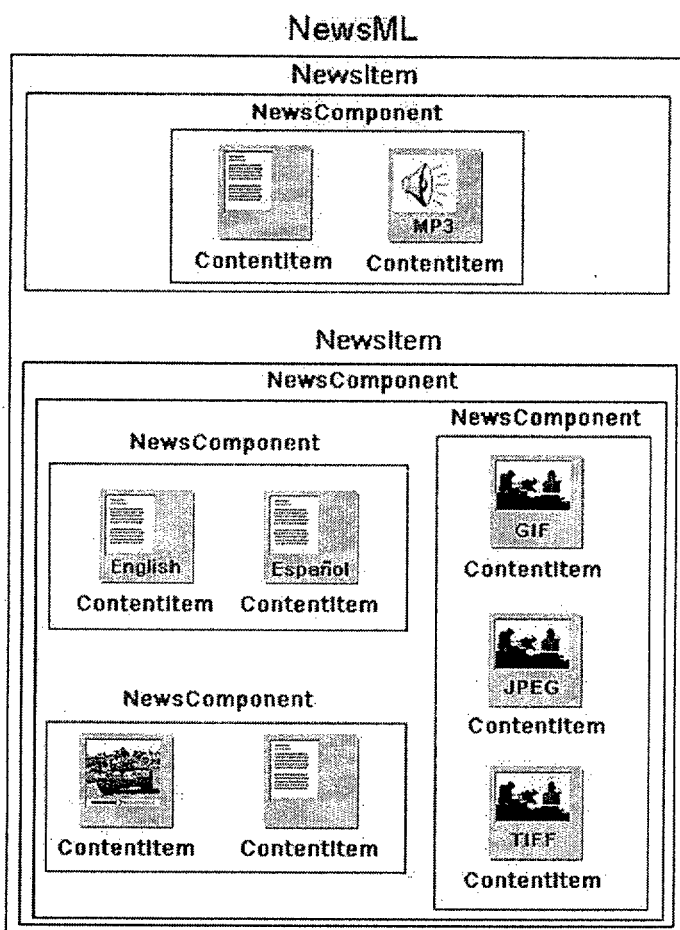


Figura 5.5.
Posible esquema con los elementos de un documento NewsML

Un documento NewsML, no obstante, puede ser bastante más complejo, como podría deducirse, por ejemplo, de la figura 5.3. Fundamentalmente, el resto de un documento NewsML puede estar formado por todo tipo de metadatos, incluidos habitualmente en toda una serie de elementos, que se referirán a:

- Un `ContentItem` concreto, que contiene un objeto (incluido a su vez dentro de un elemento `DataContent`). En este caso, se referirán a detalles físicos del objeto contenido: tipo de objeto, formato, tamaño, duración, etc.

- Un `NewsComponent` concreto, es decir, un conjunto de objetos (que puede equivaler a una pieza, aunque no necesariamente) o conjunto de conjuntos de objetos. Dadas todas estas posibilidades, el conjunto de metadatos puede ser muy amplio:
 - De tipo periodístico: titular, subtítulo, autor, fecha de creación, palabras clave, serie periodística a la que pertenece el documento, etc.
 - De tipo temático: género periodístico, clasificación temática según varios niveles de especialización definidos mediante los *códigos temáticos* de IPTC, temas concretos, etc..
 - De tipo técnico, para la gestión del contenido en el flujo de trabajo de la empresa periodística: nombre del fichero que contiene el documento, código de identificación, medio productor y medio distribuidor de la noticia, persona que edita finalmente la noticia en el medio que la publica, etc.
 - De tipo legal: referido a los derechos de propiedad y de uso sobre el contenido del documento.
 - Finalmente, cualquier otro tipo de metadatos que pueda definirse a voluntad de alguno de los medios que maneja el documento.
- Un `NewsItem` o, simplificando, cualquier parte de un documento físico NewsML que pueda ser reutilizada por separado en un futuro: puede ser un documento completo (si el documento NewsML contiene más de un documento periodístico) o puede ser una pieza.

Este elemento contiene básicamente metadatos de dos tipos. El primer conjunto de los mismos sirve para identificar inequívocamente al elemento, mediante códigos para la empresa proveedora de la noticia, para el propio `NewsItem`, para la fecha de inclusión del `NewsItem` en el flujo de trabajo del medio que recibió la noticia y, si la hubiera, para indicar que el

NewsItem ha sufrido una revisión; todos estos datos deben indicarse mediante un código legible automáticamente por una aplicación informática, aunque también pueden indicarse de forma legible por una persona.

El segundo conjunto de datos es utilizado para el manejo del ítem: su fecha de creación y revisión, la indicación sobre si debe o no ser publicado, la urgencia estimada de su difusión y la relación del NewsItem con otros, entre diversas posibilidades.

- Todo un documento NewsML. En este caso, los metadatos serán códigos de transmisión telemática de ese documento: código del medio que envía la noticia y del medio que la recibe, fecha y hora de envío, y datos adicionales.

Cada uno de estos metadatos va incluido en un elemento propio, agrupados en la mayoría de los casos, a su vez, dentro de un elemento mayor. Por ejemplo: los metadatos de un NewsComponent están agrupados en elementos hijos que a su vez están desglosados en elementos “nietos” de NewsComponent. Estos elementos hijos son Newslines (metadatos periodísticos), AdministrativeMetadata (metadatos técnicos), RightsMetadata (metadatos sobre derechos), DescriptiveMetadata (metadatos temáticos y de género) y, finalmente, Metadata, que incluye cualquier otro tipo de metadatos no definido expresamente en NewsML; estos últimos pueden tener una estructura y contenidos definidos por terceros (por ejemplo, metadatos propios de un único medio).

Todos los elementos de NewsML, sea cual sea su nivel de detalle, pueden llevar, además un identificador único interno, mediante un atributo Duid, que permite localizar a ese elemento dentro del documento.

Además, cada uno de los elementos de NewsML pueden llevar un atributo Euid, que identifica a ese elemento inequívocamente respecto a cualquier otro elemento similar: dos NewsComponent hijos del mismo NewsItem deberían llevar distintos valores para el atributo Euid.

Duid y Euid son herramientas útiles para la gestión de documentos NewsML, por ejemplo, en un sistema de gestión de contenidos, permitiendo su localización inmediata por parte del sistema.

5.3.2. Mecanismos de descripción en NewsML: empleo de lenguajes documentales

El contenido de muchos de los elementos NewsML, ya sea a través de su DTD o de su esquema, está restringido expresamente en la especificación del lenguaje: por ejemplo, las menciones referentes a una fecha en los distintos metadatos que utilizan este formato deben supeditarse a la norma ISO 8601 (INTERNATIONAL STANDARD ORGANIZATION, 2000), o bien, son susceptibles de restricción mediante la introducción de términos provenientes de un lenguaje documental. Esta característica pretende que la descripción de elementos con contenido común se ajuste, en lo posible, a sistemas de clasificación previamente definidos, facilitando un mayor control técnico y documental. Para ello, NewsML utiliza dos mecanismos de asignación de términos (ALLEN, 2002, p. 4-22):

El primero de estos mecanismos es el uso del elemento `TopicSet`, que permite incluir un lenguaje documental (una lista de autoridades, una lista de encabezamiento de materias, etc.) dentro de un documento NewsML. Cada una de las entradas de la lista se incluye dentro de un elemento `Topic`, con diversos atributos o elementos que especifican la dependencia jerárquica, en su caso, de la entrada respecto a otros elementos del `TopicSet`, asignándole además un código concreto que será usado al referirse a ese `Topic` desde otro elemento del documento NewsML. El elemento `TopicSet` se coloca al principio del elemento raíz NewsML, de un elemento `NewsItem` o de un elemento `NewsComponent`, y puede tomar datos a su vez de otros `TopicSet` situados en otros elementos de un documento NewsML.

```
<NewsML>
...
  <TopicSet ...>
    <Topic...>
    <Topic...>
  </TopicSet>
```

```
<NewsItem>
  <NewsManagement>
    ...
  </NewsManagement>
  <NewsComponent>
    <TopicSet ...>
      <Topic...>
        <Topic...>
      </TopicSet>
    </NewsComponent>
  </NewsItem>
</NewsML>
```

Figura 5.6.
Algunas posibles ubicaciones de un elemento TopicSet

El segundo mecanismo incluye el uso del elemento `Catalog`, que cumple a su vez dos funciones. La primera función consiste en indicar dónde se encuentra un lenguaje documental (generalmente, si es externo al documento) usado por diversos elementos de metadatos en un documento NewsML; `Catalog` apunta físicamente al fichero que contiene ese lenguaje documental (si es externo) o al `TopicSet` concreto (si es interno) mediante una URL o un URN, incluido en el elemento hijo `Resource`.

La segunda función de `Catalog` reside en indicar, si es preciso, a qué elementos se les aplica por defecto un determinado lenguaje documental, mediante el elemento `DefaultVocabularyFor`. El contenido de este elemento es una ruta, mediante notación XPath (CLARK y DeROSE, 1999), para localizar ese elemento dentro del documento NewsML. Puede haber varios `DefaultVocabularyFor` en el mismo `Catalog`. También puede haber varios elementos `Catalog` en el mismo documento NewsML, en cuyo caso a un elemento siempre se le aplica el lenguaje documental del elemento `Catalog` más cercano.

Un elemento `Catalog` puede ser incluido al comienzo del elemento raíz NewsML, o desde dentro de alguno de estos elementos: `NewsItem`, `NewsComponent` y sus elementos hijos de metadatos ya mencionados, `ContentItem` y los elementos `TopicSet` y `Topic`. Es decir, puede emplearse un lenguaje documental del que utilizar los descriptores empleados en todo el documento y, adicionalmente, emplear

otro lenguaje documental para una parte concreta del documento, como en la siguiente figura:

```

<NewsML>
  ...
  <Catalog>
    <Resource="plantilla.xml"/>
  </Catalog>
  <NewsItem>
    <NewsManagement>
      ...
    </NewsManagement>
    <NewsComponent>
      <Catalog>
        <Resource="colaboradores.xml"/>
      </Catalog>
      ...
    </NewsComponent>
  </NewsItem>
</NewsML>

```

Figura 5.7.

Ejemplo de un documento al que se le aplica un lenguaje documental, incluido en el fichero externo *plantilla.xml*. Además, solo en el elemento *NewsComponent* se aplica el lenguaje documental en el fichero externo *colaboradores.xml*

El ejemplo anterior podría servir para una información en la que se hace referencia a varios redactores que forman parte de la plantilla del medio que elabora esa noticia. Por ejemplo, si se trata de una noticia con varios objetos textuales, cada uno de ellos puede estar elaborado por un periodista diferente. Si se identifica dentro del documento cada periodista con un código, se especifica que todos los códigos corresponden a periodistas de plantilla, y que la equivalencia entre el código y el nombre del periodista está en el fichero **plantilla.xml**.

No obstante, uno de esos objetos puede estar elaborado por un colaborador externo al medio. En ese caso, al describir ese objeto, se especifica que el código del autor corresponde a un periodista colaborador, y se indica que la equivalencia entre ese código y el nombre del periodista está en el fichero **colaboradores.xml**.

5.3.2.1. Lenguajes documentales en NewsML: el uso de clasificaciones IPTC

En lo que se refiere a los contenidos de estos metadatos, algunos de ellos deben incluir necesariamente un término tomado de una clasificación definida por el IPTC. Los elementos que incluyen este contenido restringido pueden ser obligatorios en NewsML o no. En el caso de los elementos obligatorios, la restricción se debe a motivos técnicos: todos ellos sirven para el control de un `NewsItem` por parte de un sistema de gestión de contenidos, y permiten el intercambio entre distintos medios. Se trata de `NewsItemType` (tipo físico de ítem: noticia, datos o mensajes internos entre proveedor y receptor), `Status` (indicación de si el `NewsItem` es publicable o no), `Priority` (grado de urgencia/importancia recomendado para la difusión del `NewsItem`) y `Location` (indicación geográfica sobre el `NewsItem`, que puede ser de diversos tipos: lugar del hecho, lugar del medio que lo produce, lugar del envío, etc.)

Existe también una serie de elementos optativos, si bien, en caso de aparecer, también deben contener datos predefinidos en lenguajes documentales del propio IPTC: estos datos irán incluidos en el elemento `SubjectCode`, que es hijo de `DescriptiveMetadata` (a su vez, hijo de un `NewsComponent`). Se trata de elementos con metadatos temáticos tomados de los *códigos temáticos* definidos por el IPTC.

Estos códigos temáticos se incluyen en uno de estos tres elementos:

- `Subject`, que establece una categoría principal. IPTC ha definido 17 categorías de este tipo, que pretenden abarcar cualquier tipo de tema, como “Economía, Negocios y Finanzas”. En este caso, le correspondería el código numérico 04000000.

- SubjectMatter, que designa una materia general derivada de la anterior, como “Servicios financieros y comerciales”, a la que le corresponde el código 04006000.
- SubjectDetail, que establece una subcategoría de la anterior, como “Servicios bancarios”, a la que le correspondería el código 04006002.
- Opcionalmente, SubjectQualifier, que establece una característica genérica aplicable a uno de los tres anteriores. Por ejemplo, si se habla de una modalidad de un deporte concreto, como la prueba de natación de 100 metros estilos, se le asignaría al contenido descrito un SubjectDetail con valor “15062019”; y para indicar que se trata de la prueba femenina, se le añadiría además un SubjectQualifier con valor “15000002”.

Obsérvese que se puede deducir cuál es el nivel o niveles anteriores de una de estas categorías por los dígitos que incluye. Todos los códigos son de ocho dígitos: los dos primeros corresponden a la categoría del Subject, los tres siguientes para SubjectMatter, y los tres últimos, para SubjectDetail. La figura siguiente muestra un ejemplo desglosado del SubjectDetail “04006002”:

Código	04	006	002
Categoría	Subject (Economía, Negocios y Finanzas)	Subject Matter (Servicios financieros y comerciales)	Subject Detail (Servicios bancarios)

Figura 5.8.
Desglose del SubjectDetail “04006002” (Servicios bancarios),
y de las categorías en las que se incluye

Como puede verse, los códigos temáticos IPTC son una lista de encabezamiento de materias, que aún continúa desarrollándose en el seno del consorcio. En un documento debe usarse solo el código de uno de los tres elementos Subject o SubjectMatter o SubjectDetail, siendo opcional utilizar uno o más SubjectQualifier.

En el **Anexo 6**, se ofrece una muestra sobre una de las categorías genéricas, todas sus subcategorías definidas por IPTC y los correspondientes códigos temáticos.

5.4. Rasgos documentales de NewsML

Documentalmente, por lo tanto, disponemos de dos utilidades para el lenguaje NewsML (sin agotar otras que salen del alcance del presente trabajo): por un lado, la inclusión o referencia de objetos y piezas, en documentos contenedores que pueden incluir ciertas relaciones entre esos objetos y piezas, y que además permiten la asignación, para cada una de esas partes, de información adjunta mediante metadatos que pueden contener términos en lenguaje libre o en lenguaje controlado.

Una segunda utilidad de NewsML radica en ser un mecanismo para la definición de lenguajes documentales, de modo que estos lenguajes contengan, mediante distintos grados de profundidad, los posibles valores que puedan contener a su vez los metadatos de otros documentos en NewsML. Asimismo, estos documentos NewsML que contienen lenguajes documentales pueden ser empleados para definir contenidos de metadatos empleados en SportsML.

Analizaremos en primer lugar el papel estructural de NewsML, es decir, su capacidad para insertar o referir objetos y piezas. En un segundo epígrafe, desarrollaremos sendas explicaciones sobre el modo de describir el contenido de estas partes, así como el papel de NewsML como mecanismo para definir lenguajes documentales.

5.4.1. Definición de estructuras documentales

5.4.1.1. La disyuntiva entre inclusión y referencia

NewsML permite incluir al menos, como unidad documental mínima, una pieza (en el elemento `NewsItem`) formada a su vez por un contenedor (en el elemento

NewsComponent) que incluye un objeto (una pieza de texto, una foto, un fichero multimedia, etc.) inserto, como se indica en la figura 5.9.

```

<NewsML>
  ...
  <NewsItem>
    ...
    <NewsComponent>
      <ContentItem>
        <DataContent>
          <!-- Aquí, el contenido del objeto -->
        </DataContent>
      </ContentItem>
    </NewsComponent>
  </NewsItem>
</NewsML>

```

Figura 5.9.
Esquema básico de un documento NewsML que contiene un objeto inserto

Otra posibilidad facilita que el objeto en cuestión no esté incluido físicamente dentro del documento NewsML, sino que esté referido; es decir, que el documento NewsML sea meramente un contenedor de metadatos referidos a uno o varios objetos concretos identificados por la dirección de los mismos, mediante un URL, como se muestra en la figura 5.10.

```

<NewsML>
  ...
  <NewsItem>
    ...
    <NewsComponent>
      <ContentItem>
        <DataContent Href="objeto1.xml">
      </ContentItem>
    </NewsComponent >
  </NewsItem >
</NewsML>

```

Figura 5.10.
Esquema básico de un documento NewsML que contiene un objeto referido.

Ambos casos utilizan finalmente un elemento `DataContent` para definir el objeto. El uso de una u otra opción contiene, desde un punto de vista documental, semejanzas y diferencias. Las semejanzas entre ambas opciones son básicamente dos: una, los metadatos que se pueden asignar en uno u otro caso son básicamente los mismos; dos, si el objeto incluye un marcado añadido en un lenguaje distinto a

NewsML (por ejemplo, si se trata de una pieza basada en NITF), ese marcado puede ser el mismo independientemente de si el objeto se encuentra físicamente dentro del documento NewsML o si se encuentra en un fichero diferente.

En cuanto a las diferencias entre ambas opciones, radican fundamentalmente en dos aspectos:

- En el caso de que el objeto sea referido, la posibilidad de que la referencia sobre la dirección del mismo no coincida con la ubicación real del objeto: bien por un error en la referencia dentro del documento NewsML, bien porque ese objeto haya sido posteriormente desplazado de su ubicación original, o bien porque ese objeto haya sido eliminado sin aviso previo. Para evitar en lo posible esta última alternativa, NewsML dispone de mecanismos para indicar si un objeto ha sido revisado, dónde se ubica el documento que contiene el historial de todas esas revisiones, y, para cada objeto, si tiene una fecha de caducidad en su uso por parte de un medio.
- En el caso de que el objeto sea incluido, debe considerarse el tamaño que el documento NewsML puede alcanzar, si ocurren uno de estos tres casos: un alto número de objetos insertados en el mismo documento NewsML, una serie de objetos que ocupen un gran tamaño, o una combinación de ambas posibilidades. Este tamaño puede plantear interrogantes sobre la necesidad de manejo directo de todos esos objetos, o bien, si resulta más útil disponer sólo del mayor número posible de metadatos sobre los mismos, dejando su obtención directa sólo para aquellos casos en que sea realmente necesario.

Un caso práctico: si una agencia ha producido 100 objetos periodísticos, puede enviar solo los metadatos de los mismos, en formato NewsML, a un medio, que puede elegir únicamente una decena de los mismos, aquellos que finalmente obtendrá. Esta opción, sin embargo, supone que, en caso de que el objeto referido contenga a su vez marcado propio (por ejemplo, que sea un texto NITF), se desprecie la información recogida en este último

marcado, dejando la descripción solamente en manos de los metadatos NewsML añadidos.

5.4.1.2. El concepto de pieza y el concepto de objeto en NewsML

El mecanismo citado en el apartado anterior es, no obstante, solo uno de los posibles que pueden utilizarse en NewsML. De hecho, en el presente trabajo se ha realizado hasta ahora una asociación, implícita o explícita, entre los elementos básicos de NewsML y sus contenidos, como la que se muestra en la siguiente figura:

Elemento	Contenido
DataContent	Un objeto periodístico
ContentItem	Contenedor de un objeto y metadatos físicos asociados
NewsComponent	Contenedor de que puede incluir varios objetos (cada uno, dentro de un DataContent, en un ContentItem) o varios contenedores de objetos
NewsItem	Información periodística que incluye un NewsComponent
NewsML	Conjunto de una o varias piezas

Figura 5.11. Asociación básica entre elementos NewsML y sus contenidos

Este esquema es ampliable. Un elemento NewsComponent puede contener, entre otras opciones:

- Un objeto, insertado o referido, mediante el elemento DataContent, dentro de ContentItem como ya se ha mencionado.

- Más de un objeto, mediante sendos elementos `ContentItem`.
- Uno o varios `NewsComponent`, insertados dentro del propio `NewsComponent`.
- Uno o varios `NewsItem`, insertados dentro del propio `NewsComponent`.
- Uno o varios `NewsItem` situados en diferentes documentos `NewsML`. En este caso, en el documento `NewsML` original, se hacen referencias a los `NewsItem` externos, mediante sendos elementos `NewsItemRef`.

Es decir: el elemento `NewsComponent` no es en realidad solo un contenedor de objetos, sino que es un contenedor que puede incluir un solo objeto, varias versiones de un mismo objeto o de una misma pieza, piezas enteras, e incluso noticias enteras, y todas las combinaciones de las anteriores, lo cual permite que un documento `NewsML` almacene prácticamente cualquier combinación informativa, en un solo documento.

Esta consideración debe ser tenida en cuenta a la hora de plantear el modo con el que se describe cada `NewsComponent`. ¿Cómo distinguir cuándo contiene objetos y cuánto contiene piezas? ¿Un contenedor con varias versiones de un mismo objeto podría ser entonces considerado como una pieza? La reflexión sobre estas preguntas se realizará en el capítulo 6, al plantearse un modelo de aplicación práctica de marcado de texto.

5.4.2. Descripción semántica de contenidos en NewsML

La otra gran utilidad de `NewsML`, además de definir estructuras de objetos y piezas combinados de diversas formas, consiste en la capacidad de añadir descripción semántica y técnica de cada una de esas partes, y de los conjuntos resultantes, mediante metadatos. En este sentido, deben hacerse dos claras distinciones:

- Por una parte, un documento NewsML puede **aplicar** metadatos a contenidos adjuntos o referidos de cualquiera de las formas ya especificadas. Esos metadatos pueden contener términos en lenguaje libre o en lenguaje controlado.
- Por otro lado, un documento NewsML puede **definir** también un lenguaje documental con términos que puedan emplearse en metadatos incluidos dentro de documentos NewsML. Los lenguajes documentales situados en diferentes ficheros pueden combinarse entre sí, mediante el uso de los citados elementos `Catalog` y `TopicSet`, entre otros, lo cual permite aplicar a esos metadatos valores con diversos grados de profundidad. Esos lenguajes documentales pueden estar incluidos como parte de un documento NewsML que además contenga o refiera piezas y objetos; o bien, pueden formar parte de un documento NewsML dedicado únicamente a definir un lenguaje documental. Un ejemplo de lenguaje documental, definido en este caso por el propio IPTC, es el que contiene los códigos temáticos (*Subject Codes*, cf. § 5.3.2.1.).

A continuación, se explicarán los mecanismos empleados por NewsML para *definir* esos lenguajes documentales, en primer lugar, y posteriormente los modos para *aplicar* los correspondientes metadatos en esos o en otros documentos NewsML, tanto con esos lenguajes documentales como mediante texto libre.

5.4.2.1. Definición de mecanismos para contener lenguajes documentales mediante el formato NewsML

Todo lenguaje documental cuya definición se incluye físicamente dentro de un documento NewsML (en adelante, se dará por sentado que es así a la hora de hablar de un “lenguaje documental”) se define básicamente mediante términos, incluidos en sendos elementos `Topic`, incluidos a su vez en listas de términos, o `TopicSet`.

Cada Topic contiene a su vez una serie de elementos: obligatoriamente, un TopicType, o tipo de término (por ejemplo: *Provincia*) y un FormalName, o nombre concreto del Topic. Un ejemplo de un TopicSet, en este caso con un único término, podría ser el de la siguiente figura.

```
<TopicSet FormalName="Provincia">
  <Topic>
    <TopicType FormalName="Provincia" />
    <FormalName>Navarra</FormalName>
  </Topic>
</TopicSet>
```

Figura 5.12.
Ejemplo básico de un elemento Topic

En la figura anterior, se observa que el Topic contiene el término *Navarra*, y que se trata de una *provincia*. Pero no se indica a qué país pertenece esa provincia. Podríamos entonces utilizar otro TopicSet como éste:

```
<TopicSet FormalName="LugarGeografico">
  <Topic>
    <TopicType FormalName="País"/>
    <FormalName Scheme="Europa">España</FormalName>
  </Topic>
</TopicSet>
```

Figura 5.13.
TopicSet incluyendo algunos atributos FormalName y Scheme

La clasificación de esta figura es superior a la empleada en la figura anterior. Obsérvese que se define un valor para la clasificación conjunta, *LugarGeográfico*, que incluye en este caso un único tipo de contenidos, *País*, y, en este caso, un único valor, *España*. Para utilizar esta clasificación dentro de la anterior, es decir, para indicar que *Navarra* es una provincia perteneciente a *España*, modificaríamos el ejemplo de la figura 3.16. de modo que quedara así:

```
<TopicSet Duid="#prov" FormalName="España" Scheme="Europa">
  <Topic>
    <TopicType FormalName="España" Scheme="Europa" />
    <FormalName Scheme="Provincia">Navarra</FormalName>
  </Topic>
</TopicSet>
```

Figura 5.14.
Topic que toma valores de otro Topic situado en un TopicSet distinto

Obsérvese que, en este caso, tanto el `TopicSet` como el `TopicType` indican que se trata de una clasificación basada en un `Topic`, denominado *España*, que pertenece a otra categoría (o `Scheme`, en la terminología usada por NewsML) llamada *Europa*. Posteriormente, se indica que el término a describir pertenece a una nueva subcategoría, llamada *Provincia*, y que el término en cuestión se denomina *Navarra*.

Las posibilidades van mucho más allá del ejemplo. Un conjunto de términos designado por `TopicSet`, como se ha indicado, puede contener diferentes categorías o `Scheme`, (continuando por el ejemplo, una lista de lugares geográficos podría incluir una lista de ciudades, otra de provincias, otra de países, otra de continentes...), relacionadas entre sí (empleando los mismos valores del contenido de los `FormalName` que definen cada categoría principal como valores del correspondiente atributo `Scheme` del `TopicType` de todos los elementos de una categoría secundaria).

Además, un `TopicSet` puede estar situado en el mismo documento que otro `TopicSet` que emplee términos del primero (definidos mediante sendos `Topic`), o bien, ambos pueden estar en documentos distintos. En este caso, el segundo `TopicSet` incluirá un atributo `Vocabulary`, con la dirección (en forma de URL o URN definido por NewsML, así como la cadena formada por el carácter “#” seguido del atributo `Duid` del `TopicSet` en el que está el `Topic`).

Como puede intuirse, la relación entre los términos incluidos en dos `Topic` distintos no tiene por qué ser jerárquica, sino que puede ser de cualquier otro tipo. Un ejemplo sería el uso de un calificador, como el código correspondiente al término *mujeres*, aplicado a un `Topic` que contuviera el término *baloncesto*, para designar que un documento o una parte del mismo se refiere en concreto al baloncesto femenino. El calificador *mujeres* podría ser uno de los previamente definidos por el IPTC mediante la clasificación denominada *SubjectQualifier* (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003b).

No obstante, el contenido de un `Topic` no tiene por qué estar formalmente basado en una categoría anterior. Una posible variante es que el `Topic` contenga uno o varios elementos `Description`, que contengan texto libre. Este último elemento puede llevar un atributo, también mediante texto libre, que explique el contenido del elemento.

```
<Topic>
  <TopicType FormalName="Persona">
    <Description Variant="Nombre">Ana Piera
  </Description>
    <Description Variant="Profesion">Escritora
  </Description>
</Topic>
```

Figura 5.15.

Ejemplo de `Topic` con contenido libre, sin definición previa por un lenguaje documental.

5.4.2.2. Aplicación de metadatos en NewsML

Los mecanismos concretos, empleados para la asignación de valores en forma de términos y códigos a cada uno de los elementos de NewsML que pueden contener metadatos son muy variados, tanto como los propios metadatos empleados. Dos ejemplos: puede ser un dato de normalización técnica, como la indicación del medio productor, mediante un código interno de control usado por el medio receptor para registrar qué noticias ha recibido de cada uno de sus proveedores (las agencias o los agregadores); o puede ser el contenido del subtítulo de una noticia, que sea añadido expresamente en NewsML para complementar su falta dentro del propio objeto.

No obstante, estos mecanismos de asignación se pueden englobar en diversas categorías, atendiendo especialmente a que el uso de elementos relativamente comunes (por ejemplo, el título y el subtítulo de un objeto informativo) contienen estructuras más o menos similares. Estas similitudes, en otros casos, dependen más del tipo de metadato que de su ubicación. Un ejemplo de este último caso son los metadatos que contienen fechas (de creación del objeto, de envío del medio productor o distribuidor a otro medio, de caducidad de los derechos de uso del documento, etc.), ajustados en la norma ISO 8601.

Basándonos en estas coincidencias, señalamos a continuación los mecanismos básicos de asignación de metadatos que pueden emplearse en un documento NewsML.

5.4.2.2.1. **Aplicación de metadatos que contengan atributos con valores basados en lenguajes documentales**

La opción aplicada a un mayor número de elementos consiste en la asignación a los mismos de un atributo `FormalName`, que contenga el nombre del término añadido en un elemento `FormalName` perteneciente a un `Topic`, como se indicó en el apartado anterior; así como la inclusión de un atributo adicional `Scheme`, que contenga el valor de una categoría contenido en el atributo `Scheme` del mismo elemento `FormalName` referido. En su caso, si además ese elemento `FormalName` está en otro documento NewsML externo, se hará referencia al lugar exacto del documento, y del `TopicSet` dentro del mismo, mediante un atributo `Vocabulary`, como se indicó también en el apartado anterior.

Un ejemplo, mediante el que se indicaría que el contenido de un documento NewsML forma parte de un servicio de noticias exclusivo para medios radicados en *Navarra*, utilizando para ello el lenguaje documental definido en la figura 3.18. (contenido en el fichero *provincias.xml*), sería el siguiente:

```
<NewsML>
  <NewsEnvelope>
    ...
    <NewsService Scheme="Provincia"
      FormalName="Navarra"
      Vocabulary="provincias.xml#prov" />
  </NewsEnvelope>
  ...
</NewsML>
```

Figura 5.16.
Ejemplo de atributo `FormalName` para asignar un término de un lenguaje documental controlado en un elemento de un documento NewsML

Esta opción se usa especialmente, aunque no en exclusiva, para aquellos elementos que pueden contener valores ya incluidos en alguno de los lenguajes documentales o clasificaciones definidas por el propio IPTC, que se explicaron en el capítulo 2. La

veintena de elementos que responden a este perfil designan tanto aspectos temáticos como relativos al flujo de trabajo del medio. Fundamentalmente, se trata de elementos que:

- Designan datos técnicos: el formato de un objeto, el tipo MIME (IANA, 2002), el tipo de objeto (texto, fotografía, dibujo, sonido, etc.) y similares.
- Definen cualidades periodísticas, como el género periodístico de un objeto, el tema tratado –mediante diversos clasificadores temáticos–, el grado de relevancia y de urgencia en su difusión, etc.
- Añaden detalles necesarios para el manejo de cada objeto o pieza en el flujo de trabajo: el tipo de público potencialmente interesado en la noticia, el papel o ubicación del objeto o de la pieza en el conjunto del medio –por ejemplo, que se trata de una columna, de un despiece, etc.–, el idioma de un objeto, el tipo de pieza, el tipo de titulares que se incluye, el producto y servicio a medida al que pertenece el envío y recepción de un documento NewsML, etc.

5.4.2.2. Asignación de metadatos mediante propiedades y valores: el elemento `Property`

Una segunda posibilidad, más precisa, es el uso del elemento `Property`, que también permite asignar términos de lenguajes documentales referidos a ciertos documentos. `Property` se utiliza como un elemento hijo que se añade al elemento que se quiere describir.

`Property` se distingue, en primer lugar, por permitir que un metadato pueda contener una descripción mediante texto libre insertada directamente al elaborar el documento, y no solo en la elaboración del lenguaje documental (como se señalaba con el uso del elemento `Description`).

Además, `Property` hace posible la asignación de metadatos mediante la estructura **propiedad-valor**. En otras palabras, `Property` facilita, además de que el valor de

un metadato esté normalizado, que también se pueda formalizar el tipo de metadato, o **propiedad**, mediante cualquier lenguaje documental. En todo caso, ninguno de los dos datos debe estar obligatoriamente normalizado.

Un caso concreto: si se pretende indicar que la persona que ha creado un objeto de texto se llama *Juan Ramírez* y es el *subdirector* del medio en el que se ha elaborado, el término *subdirector* sería entendido como una **propiedad** de ese objeto (el cargo de su redactor dentro del medio), mientras que *Juan Ramírez* sería un **valor** de esa propiedad (sin descartar que pudiera haber más valores posibles de esa propiedad o, siguiendo el ejemplo, otros subdirectores de ese medio).

El elemento `Property` permite que tanto la propiedad (mediante un atributo `FormalName`) como el valor (mediante un atributo `ValueRef`, que remite a un código o término restringidos en un lenguaje documental, así como un atributo `Value`, que equivale al anterior pero en lenguaje libre) se basen en un lenguaje documental. Además, permite el uso de atributos para designar la categoría en la que se engloba la propiedad (mediante el atributo `Scheme`) y el valor (mediante el atributo `AllowedScheme`). También se permite designar el documento que contiene el `TopicSet` con las propiedades, mediante el ya citado atributo `Vocabulary`, y el documento que contiene el `TopicSet` en el que se incluyen los valores, con el atributo `AllowedValues`. La siguiente figura muestra un ejemplo completo de la aplicación de un elemento `Property`.

```
<AdministrativeMetadata>
  <Creator>
    <Party>
      <Property FormalName="subdirector"
        Scheme="directivos"
        Vocabulary="staff.xml#principales" ValueRef="234"
        Value="Juan Ramírez" AllowedScheme="redacción"
        AllowedValues="personal.xml" />
    </Party>
  </Creator>
</AdministrativeMetadata>
```

Figura 5.17.

Ejemplo de un elemento `Property` empleado para designar al autor de una información

En este caso, se trataría de un elemento para definir al autor de una información. En primer lugar, se definen los datos relativos a la propiedad del autor: en este caso, se trata de un *subdirector*, término que forma parte de la categoría *directivos*, incluida en un fichero externo (denominado *staff.xml*, dentro del cual habría un *TopicSet* cuyo atributo *FormalName* sería *principales*) que incluye a todos los responsables de alguna sección en el medio.

Posteriormente, se añaden los datos para designar el valor de la propiedad, es decir, el nombre de la persona, en este caso: siguiendo con el ejemplo anterior, el citado *Juan Ramírez*, referido además por el código *234*, con el que está registrado en el fichero de personas que trabajan en el medio (dando por hecho que todos los trabajadores del mismo tienen un código que les distingue inequívocamente de cualquier otro miembro del personal con el mismo nombre y apellidos), *personal.xml* y, dentro de él, en la categoría de *redacción* (formada imaginariamente por las personas del periódico que trabajan en la redacción, diferenciándolas de las que trabajan en otras áreas, como la de publicidad, por ejemplo).

Property permite además incluir en su seno otros elementos *Property*, de modo que un conjunto propiedad-valor pueda ser además a su vez propiedad de otro valor, con tantos niveles de profundidad como sea preciso. Por ejemplo: que una noticia tenga como protagonista (propiedad) a una cantante, Luz Casal (valor); y que la cantante Luz Casal (nueva propiedad conjunta) tenga 44 años (valor)...

Este elemento sirve para añadir todo tipo de metadatos complejos a un *NewsComponent* (incluyéndose a su vez dentro de sus elementos contenedores de metadatos *AdministrativeMetadata*, *DescriptiveMetadata*, *RightsMetadata* y *Metadata*), así como a términos incluidos en elementos *Topic*, *Characteristics* (que describe datos físicos, como el formato o la duración, de un *ContentItem*), *Party* (que sirve para designar un nombre de persona en concreto a un participante en el proceso de elaboración de un documento NewsML: el autor, el corrector, etc., como se señala en la figura anterior) y *Location* (que indica un lugar y su relación con el hecho: el lugar en el que sucede



la noticia, o en el que se elabora el objeto, o el lugar al que afecta una información (por ejemplo, si una noticia se refiere a todo un país, como España), etc.

5.4.2.2.3. *Indicación física de la ubicación del lenguaje documental*

Tanto en el uso de `Property` como en el uso de atributos con contenidos previamente definidos en un lenguaje documental, la definición de la ubicación física en la que se encuentra el lenguaje documental puede realizarse de dos maneras: con el ya citado atributo `Vocabulary`, o bien, mediante un elemento previo `Catalog` que indique cuál es el fichero que contiene ese lenguaje documental (o vocabulario, en la terminología NewsML) y sobre qué elemento y atributos se aplica.

El ejemplo de la siguiente figura es una modificación de la expuesta en la figura 5.17, sustituyendo el atributo `Vocabulary` por el citado elemento `Catalog`, con sus subelementos correspondientes.

```

<Catalog>
  <Resource Duid="cargos">
    <Url>staff.xml#principales</Url>
    <DefaultVocabularyFor Context="Party@ValueRef" />
  </Resource>
  ...
</Catalog>
...
<AdministrativeMetadata>
  <Creator>
    <Party>
      <Property FormalName="subdirector"
        Scheme="directivos"
        ValueRef="234"
        Value="Juan Ramírez" AllowedScheme="redacción"
        AllowedValues="personal.xml"
      />
    </Party>
  </Creator>
</AdministrativeMetadata>

```

Figura 5.18.

Ejemplo de un elemento `Catalog` para definir el lenguaje documental empleado en otro elemento de un documento NewsML

El caso mostrado en la figura anterior muestra como el elemento `Catalog` define el fichero que contiene el lenguaje documental a emplear (la categoría *principales* en

staff.xml) y dónde se aplica (en el atributo ValueRef del elemento Property que sea hijo del elemento Party).

5.4.2.2.4. Referencia a lenguajes documentales mediante punteros

Una alternativa, para ciertos elementos, reside en la indicación directa a otro punto del mismo documento, o de otro fichero externo, cuya información se utiliza para describir de modo formal (es decir, mediante un lenguaje documental) alguno de sus contenidos. En ambos casos, comparten el mismo funcionamiento: el uso de un atributo que apunta a una URL, a un URN, o bien, a un atributo Duid de algún elemento del mismo documento, o de otro, donde está el elemento con el citado atributo Href.

Un ejemplo es el elemento Origin, utilizado junto a ciertos metadatos de tipo periodístico, aplicados a un NewsComponent, para designar al autor de un objeto, a datos sobre colaboradores de este autor o sobre autoría múltiple, a palabras clave, al poseedor del copyright, etc. Todos estos elementos pueden contener texto libre, y/o incluir una referencia al atributo Duid de otro elemento, en ese documento o en otro cualquiera, que identifique de un modo más formal ese contenido. En ambos casos, el puntero se incluye mediante el atributo Href. Origin puede además incluir otros elementos Origin, lo cual permite, por ejemplo, que una misma referencia pueda incluir varios datos formalizados y varios sin formalizar.

```
<NewsLines>
  <Byline>
    <Origin Href="#jr" />
  </Byline>
</NewsLines>
...
<AdministrativeMetadata>
  <Creator Duid="jr">
    <Party>
      <Property FormalName="subdirector"
        Scheme="directivos"
        ValueRef="234"
        Value="Juan Ramírez" AllowedScheme="redacción"
        AllowedValues="personal.xml" />
    </Party>
  </Creator>
</AdministrativeMetadata>
```

```
</AdministrativeMetadata>
...
```

Figura 5.19.
Ejemplo básico de un elemento `Origin`

Por su parte, el elemento `Topic` puede llevar un atributo `Details` que tiene la misma función y el mismo mecanismo que el citado atributo `Href` de `Origin`: ofrecer información adicional, formalizada o no, sobre ese `Topic`.

El elemento `RevisionHistory`, que también contiene un atributo `Href`, apunta a un documento que contiene, en caso de que haya varias revisiones de un `NewsItem`, a un fichero histórico con información sobre el proceso de esas revisiones.

Otra posibilidad se abre con el empleo del atributo `DuidRef`, que permite que pueda variarse el contenido de un elemento incluido en un `NewsItem` perteneciente a otro documento `NewsML`: cambiarlo, eliminarlo o añadirle contenido antes o después del citado elemento. Para ello, el `DuidRef` hace referencia al atributo `Duid` del elemento que se quiere cambiar, mientras que se le aplicará a ese último elemento el cambio indicado por el primero.

Una posibilidad, solo válida cuando la definición de un `Topic` se realiza dentro del mismo documento en el que se insertan o refieren objetos, es el atributo también llamado `Topic`, que puede aparecer dentro de tres elementos:

- `TopicOccurrence`, incluido a su vez dentro del elemento `DescriptiveMetadata`, el contenedor con los metadatos periodísticos que forman parte de un `NewsComponent`. En este caso, el atributo `Topic` refiere al `Duid` de un elemento `Topic` que está, como indicamos, en el mismo documento. Por ejemplo:

```
<NewsML>
  <TopicSet FormalName="deportistas">
    <Topic Duid="falonso">
      <TopicType FormalName="Persona">
        <Description Variant="Nombre">Fernando Alonso
      </Description>
      <Description Variant="Profesion">Piloto de F1
    </Description>
  </TopicSet>
</NewsML>
```



```
</Topic>
</TopicSet>
...
<NewsItem>
  <NewsComponent>
    <DescriptiveMetadata>
      <TopicOccurrence Topic="#falonso">
    </DescriptiveMetadata>
    ...
  </NewsComponent>
  ...
</NewsItem>
<NewsML>
```

Figura 5.20.
Ejemplo de TopicOccurrence

- El ya citado elemento Location, también incluido dentro de DescriptiveMetadata.
- El también citado elemento Party.

5.4.2.2.5. Atributos para la identificación de autores de metadatos

Un mecanismo adicional para la asignación de metadatos permite la inclusión, mediante atributos aplicables a un buen número de elementos, de metadatos que identifican la persona que ha asignado un valor a un metadato, el grado de importancia que esa persona le da a su propia asignación, el grado de confianza con el que se ha introducido ese metadato, el modo en el que se ha asignado el metadato (por ejemplo, de modo manual, a través de una determinada herramienta, de modo automatizado, etc.) y, finalmente, la fecha y la hora en la que se ha asignado ese metadato. Este conjunto de mecanismos tiene una serie de implicaciones para el funcionamiento del medio que desarrollaremos más adelante, en el capítulo 4.

Todos los valores de estos atributos deben estar basados en un lenguaje controlado. De hecho, algunos de estos valores, como la confianza –entendida como el valor que se le da a la propia asignación de un valor por parte de su autor–, disponen ya de sendos lenguajes propios definidos por el IPTC (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003a) . En todos los casos, habrá

de hacerse referencia a ese lenguaje controlado desde un anterior elemento `Catalog`, como ya se ha señalado en un apartado anterior, excepto en el caso del atributo conjunto que indica la fecha y hora de la asignación de metadatos, la cual deberá estar basada en el citado formato ISO 8601.

Estos metadatos pueden ser asignados fundamentalmente a los elementos temáticos, a las referencias personales sobre la autoría de las informaciones, a los relativos a los derechos de uso y a otros como los que indican el género, la lengua, las referencias geográficas, la asignación de metadatos complejos mediante propiedades y valores, etc.

5.4.2.2.6. Inserción de metadatos adicionales a los ya existentes

NewsML permite crear metadatos de cualquier tipo añadidos a los ya existentes. Para ello, dispone del citado elemento `Metadata`, que descende del elemento `NewsComponent`. Mediante el oportuno uso de uno o más elementos `Property`, pueden utilizarse otras posibilidades de descripción no incluidas entre los elementos de la especificación NewsML. Un caso concreto: el precio de venta de ese documento a otros medios.

En principio, las propiedades correspondientes sí deben estar definidas, mediante sendos `Topic`, en un lenguaje documental creado con NewsML. El comité que se encarga de actualizar la especificación NewsML está barajando la posibilidad de que tanto las propiedades como los valores de los metadatos puedan ser elaborados en cualquier lenguaje como, por ejemplo, RDF o PRISM, entre otros (LE MEUR, 2003b); es decir, que puedan emplearse automáticamente clasificaciones hechas, por ejemplo, en RDF, sin necesidad de que haya que definir lenguajes documentales en documento que traduzcan la clasificación en formato RDF al formato NewsML.

5.4.2.3. Lenguajes documentales concretos definidos por IPTC: utilidades e interrogantes

Empleando los mecanismos que acabamos de explicar, el propio IPTC ha definido además toda una serie de lenguajes documentales concretos, entre los que destaca el llamado Sistema de Referencia Temática (Subject Reference System) que contiene los diversos códigos temáticos utilizados a su vez en los documentos elaborados mediante lenguajes de marcado del IPTC, tanto en NewsML como en NITF y SportsML (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL y Newspaper Association of America, 2003). Todos los documentos que forman el Sistema de Referencia Temática son documentos NewsML formados por conjuntos de `TopicSet`, cada uno de ellos con sucesivos `Topic` que contienen los términos, en diferentes niveles de profundidad.

Pero no solo se emplea esta opción para la descripción temática. Como ya se ha mencionado, IPTC ha elaborado varios lenguajes adicionales, a los que también ha denominado en sí `TopicSet` (dado que su contenido también está formado por uno o varios elementos `TopicSet`), cuyos elementos contienen los términos de IPTC usados para normalizar la mención al tipo de ítem (texto, datos brutos, etc.), el lenguaje del documento, el país o la audiencia potencialmente interesada en un documento periodístico, entre otras posibilidades. (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2003c)

IPTC también ha dado por válidas algunas listas de autoridades y encabezamientos de materias, como los códigos ISO de lengua (INTERNATIONAL STANDARD ORGANIZATION, 2002) y moneda (INTERNATIONAL STANDARD ORGANIZATION, 2001), la clasificación estadounidense NAICS de actividades industriales (U.S. CENSUS BUREAU, 2002), los códigos de empresas que forman parte del índice NASDAQ de la Bolsa de Nueva York (NASDAQ, 2003) y la clasificación de regiones geográficas desarrollada por Naciones Unidas (NACIONES UNIDAS, 2003). Todas ellas han sido adaptadas al formato NewsML por el propio IPTC, asumiéndolas como propias.

El empleo de estos tres tipos de clasificaciones es promovido por IPTC con el fin de normalizar, en lo posible, la descripción de contenidos en todo tipo de medios. Ello

supondría la posibilidad de reducir el esfuerzo documental, al menos, para la descripción básica de los contenidos.

Por ejemplo: supongamos que un conjunto de medios decidiera compartir un mismo sistema de clasificación básico (como el Sistema de Referencia Temática) con una agencia que difunda información continua a los mismos. En ese caso, los medios no solo recibirían los contenidos, sino también, las descripciones de los mismos. Si las dieran por válidas, los medios dispondrían por tanto de un doble valor: no solo esos contenidos, sino también herramientas para recuperar esos contenidos, o para demandar contenidos relacionados a la agencia en un futuro.

Sin embargo, los medios también pueden enfrentarse a un interrogante: ¿estarán aceptando, junto con las herramientas tecnológicas, una determinada forma de clasificar las noticias, impuesta de una u otra forma por la agencia, y con ella, una determinada visión del mundo? Al ser una institución promovida por buena parte de las grandes agencias en todo el mundo, los códigos temáticos IPTC pretenden ser válidos universalmente, con el fin de facilitar el intercambio entre cualquier medio. Pero no olvidemos el fuerte sesgo anglosajón de las agencias internacionales, reflejado también en el funcionamiento del propio IPTC.

Los medios de comunicación están tratando con contenidos que, como hemos descrito en el capítulo 2, son de tipo universal: pero además, presentan un fuerte grado de ideologización. De hecho, los propios centros de documentación de medios suelen afirmar que el funcionamiento en cada medio es muy distinto respecto a otros; y también lo son las clasificaciones adoptadas en cada centro, teñidas por una serie de principios ideológicos aplicados sobre un modo concreto de jerarquizar la realidad.

Esta concepción sobre el sesgo de uno u otro sistema de clasificación será, no obstante, menor en el caso de aquellas clasificaciones más técnicas, como la que designa el tipo de documento que se envía, o que emplean una lista “objetiva” de términos, elaborada por una institución aceptada como válida: por ejemplo, es el caso de la clasificación del NASDAQ, útil para identificar noticias sobre empresas que

coticen en la Bolsa neoyorquina. En estos casos, sí será útil identificar en concreto los documentos que traten de esas instituciones y de los temas relacionados, por ejemplo, con su cotización bursátil.

5.4.2.3.1. Flexibilidad en la aplicación de metadatos: atractivo para los medios

NewsML ofrece además otro posible atractivo para su implantación en los medios, al menos, desde un punto de vista documental: la enorme flexibilidad en la asignación de descripciones documentales

- Desde el punto de vista de los posibles lugares de un documento en los que se puede insertar esta descripción.
- Desde la perspectiva de los propios mecanismos empleados para definir lenguajes documentales y para aplicarlos en documentos periodísticos concretos.

Ello supone, visto de modo interno –como acabamos de describir–, que el lenguaje pueda llegar a ser complicado en cuanto a su codificación. No obstante, NewsML no ha sido definido para la elaboración manual de marcado de documentos, sino solo como una herramienta de representación interna que sea susceptible de aceptar todo tipo de interfaces entre el usuario –tanto en la elaboración de documentos NewsML como en su recuperación– y los documentos.

5.5. NITF: descripción estructural de noticias textuales

Como se ha mencionado en la sinopsis histórica sobre el marcado periodístico, NITF fue cronológicamente anterior a NewsML: sus orígenes se remontan a 1995, cuando surgió el primer prototipo conocido como NIUTF, basado en SGML y con una fuerte influencia de HTML. El advenimiento de XML hizo que se realizara una nueva

versión basada en este lenguaje en el seno del IPTC, con el nombre ya de NITF, en 1999.

Posteriormente surgieron otras iniciativas de marcado de documentos periodísticos, como NML (COVER, 1999), auspiciada por la NAA estadounidense, o más tarde el citado lenguaje XMLNews. Ambas acabaron subsumiéndose en NITF, que ha continuado evolucionando hasta la actualidad. La versión 3.1, en forma de DTD (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC), 2002a), fue hecha pública en julio de 2002, y la actual versión 3.2, en octubre de 2003.

NITF es fundamentalmente un lenguaje para el marcado de noticias textuales: aunque incorpora la posibilidad de incluir tanto contenidos multimedia, dentro de un documento NITF, como referencias a documentos multimedia externos, el conjunto de elementos de NITF está orientado a la definición de noticias textuales. NITF hace posible la descripción de su estructura, con posibilidad de llegar hasta sus partes más pequeñas, y la inclusión de todo tipo de información añadida, mediante marcado documental inserto en el texto o adjunto al mismo.

Esta vocación de lenguaje empleado para la descripción textual profunda, y su modularidad externa, es decir, la capacidad de combinar su uso con el de otros lenguajes, han hecho que el propio IPTC sugiera el uso de NITF como el lenguaje con el que se realicen los contenidos textuales insertos en otros lenguajes de la familia, como el citado NewsML o, según se describe más adelante, SportsML. Pero, además de esa modularidad externa, NITF presenta una marcada modularidad interna: por una parte, al ser un lenguaje marcadamente estructural, permite detallar desde las estructuras de informaciones más básicas (por ejemplo: un mero párrafo de texto) hasta las más complejas, tanto en lo que se refiere al modo estructural (por ejemplo, múltiples noticias, cada una de ellas descompuesta a su vez en diversas piezas) como semántica (por ejemplo, distinguir cada una de las menciones a un detalle que pueda responder a las preguntas consideradas como clásicas en el periodismo: quién, qué, cuándo, cómo, dónde y, aunque en menor medida, por qué).

Estas capacidades son factibles mediante toda una red de elementos que, asimismo, incluyen el uso de lenguajes documentales, controlados o no, mediante una serie de atributos y, en muchos casos, valores restringidos, ya sean los definidos por IPTC (códigos temáticos, Sistema de Referencia Temática), etc.) o los definidos por otros medios o instituciones concretas.

Por lo tanto, las tres peculiaridades básicas de NITF, estructuración, modularidad y capacidad de descripción, son las que nos permiten describir desde un punto de vista documental las utilidades del lenguaje.

5.5.1. Tipos de elementos: obligatorios y opcionales

NITF es un lenguaje estructural, que permite delimitar cada uno de los componentes de un documento periodístico fundamentalmente textual.

No obstante, esta estructura, en origen, está basada en HTML. La existencia de elementos como, por citar algunos, `body`, para designar al cuerpo de la noticia, `p`, para designar al párrafo, `table`, para marcar una tabla, o la distinción entre listas ordenadas, con `ol`, y desordenadas, con `ul`, ratifican este hecho; más aún, la indefinición estructural en algunos momentos, con la inclusión de elementos más bien destinados a la distinción tipográfica, como `pre` o `em`, corroboran esta observación, apostillada con el hecho de que, entre las listas de correo en las que se discute la evolución actual del lenguaje, XHTML aparece mencionado en algunos momentos como fuente de inspiración para el futuro (LE MEUR, 2003a).

Sin embargo, no es correcto deducir de todo ello que NITF sea igual que HTML. Su capacidad de descripción estructural, y con ello, su utilidad documental, son mucho mayores. En primer lugar, las posibles estructuras de NITF son independientes de su representación futura, ya sea visual o auditiva. Por otra parte, como se explicará más adelante, esta estructura es bastante flexible, ofreciendo numerosas posibilidades de uso y combinación de elementos, muchos de los cuales pueden estar situados en diferentes partes de la estructura.

Un documento básico de NITF llevará un elemento raíz, denominado `nitf`, del cual colgará necesariamente un elemento `body` y, dentro de él, algunas, varias o todas estas partes que, en cualquier caso, deberán seguir el siguiente orden de aparición: `body.head`, `body.content` y `body.end`.

`body.head` contiene tanto elementos que pueden formar parte del contenido expreso de un documento periodístico, ya sean los titulares, la autoría, etc., o bien, metadatos añadidos, como un resumen del documento, datos relativos a los derechos, a la empresa distribuidora de las noticias, la fecha, la serie de artículos a la que pertenece el documento NITF, etc.

`body.content` incluye en su seno la parte central del documento. Puede agrupar, a su vez, algunas, varias o todas estas partes: una sección inicial, que puede contener diversos metadatos o partes insertas en el contenido, de las cuales unas también pueden estar presentes en `body.head`, como la autoría, la fecha o el resumen, y otras solo pueden situarse al comienzo de `body.content`, como `tobject`, que contiene sendos atributos con los *códigos temáticos* IPTC, o metadatos para la inclusión de descriptores tanto libres como controlados, como `classifier` o `key-list`, así como metadatos que contienen información sobre el copyright.

La segunda parte de `body.content` es la que incluye la mayoría del contenido de un documento periodístico, mediante elementos que incluyen párrafos, tablas –tanto en el formato definido por NITF como en formatos definidos por un usuario particular, como podría ser un medio determinado–, listas, citas, objetos multimedia –mediante el elemento `media`, que incluye otros elementos hijos–, subtítulos y otros elementos menores. En cuanto a la última parte de `body.content`, se incluye la mención a personas que han colaborado en la redacción de la pieza incluida en este elemento.

Por último, `body.end` está formado por datos de autores adicionales del documento NITF (por ejemplo, periodistas que han colaborado en el conjunto del documento, y no solo en una parte), y bibliografía complementaria, si la hubiera.

`body` es uno de los dos elementos básicos de un documento NITF. Además, al igual que ocurre en un documento HTML, existe un elemento previo opcional, denominado `head`, que contiene metadatos, algunos de los cuales ya han sido mencionados dentro de otros elementos pertenecientes al `body`.

Básicamente, estos metadatos pueden ser: un título para el documento NITF (que no tiene por qué coincidir con un titular); uno o varios códigos temáticos IPTC, también mediante `tobject`; un código equivalente al anterior, basado en la clasificación temática utilizada en las primeras versiones del citado IIM; un conjunto de metadatos técnicos (código de identificación del documento; grado de urgencia; fecha de elaboración del documento, de publicación y de caducidad, si la tuviera; etc.) y temáticos (palabras clave en lenguaje libre, identificadores concretos de personas, lugares, cargos, hechos, etc.); una serie de metadatos físicos (datos sobre publicación, en el caso de ser un documento impreso, o datos de formato y reproducción, en el caso de documentos con contenidos multimedia); datos sobre personas que hayan modificado el documento original (por ejemplo, redactores jefe); y además, cualquier otro metadato definible por un medio concreto, a su libre voluntad.

Estos metadatos pueden contener datos en lenguaje libre, en algunos casos, o tener un contenido restringido mediante opciones definidas en el propio lenguaje NITF, como el grado de urgencia, cuyo valor consiste en un dígito: a mayor valor, menor urgencia. En otros casos, como indicamos con `tobject`, los valores están definidos en lenguajes documentales creados por el propio IPTC.

En síntesis, un posible esquema con los elementos principales de NITF podría ser el siguiente:

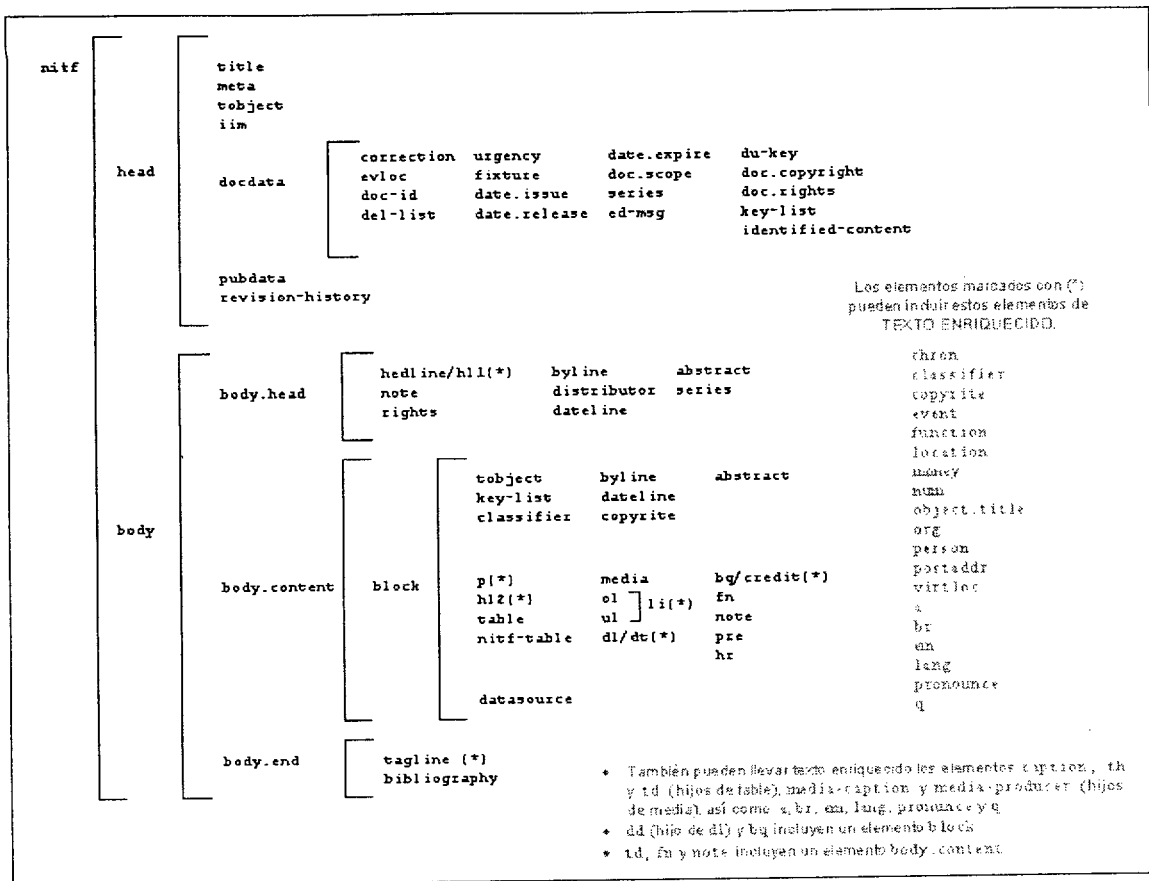


Figura 5.21. Esquema básico de elementos de NITF

5.5.2. Combinación de los elementos en NITF mediante módulos flexibles

Un documento NITF es, desde el punto de vista estructural, un conjunto de módulos que permiten la definición de estructuras, las cuales pueden ser muy sencillas, con documentos periodísticos formados por una sola línea, o muy complejas, como la inclusión en un mismo documento NITF de centenares de documentos periodísticos, por ejemplo, sobre un mismo tema.

Para ello, la gramática NITF, como ya se ha esbozado, ofrece flexibles combinaciones de elementos. En principio, apenas con media docena de elementos se puede establecer un documento NITF, mientras que una estructura lógica formada por varias piezas para un mismo documento periodístico, o incluso varios

documentos en conjunto, son perfectamente encajables, diferenciando claramente cada uno de estos contenidos del resto.

Esta modularidad se refiere tanto a los elementos destinados a estructurar el contenido como a los elementos dedicados a describir dicho contenido, sus cualidades o el contexto en el que se enmarcan. En el caso estructural, como se ha podido deducir, NITF permite una combinación de metadatos y estructura, mediante el empleo de `head` y `body`, de modo que esos metadatos afecten al conjunto del documento, es decir, empleando un modelo de marcado documental adjunto.

NITF ofrece también una segunda posibilidad de representación, en la que únicamente se utilice el elemento `body`, formado a su vez por una serie inicial de metadatos o de datos estructurales, una parte central que incluye el contenido documental y una parte final de añadido documental.

Pero, más aún, la parte central de `body`, el elemento `body.content`, puede a su vez disponer de otras tres posibles partes. En este último caso, la estructura es similar: una inicial con metadatos o con datos estructurales, una central que incluye únicamente elementos estructurales, y una final con algún metadato adicional. Cuando se contienen al menos dos de estas partes (la central y otra), el conjunto debe ir dentro de un bloque, definible mediante el elemento `block`, hijo de `body.content` (véase la figura 5.21.).

Esta estructura máxima de capas de elementos puede combinarse de diferentes maneras. Solo la parte central final, integrada o no en un elemento `block`, y que contiene la estructura con el contenido de un objeto periodístico, es obligatoria en un documento NITF, mientras que todas las demás partes cumplen dos funciones: permitir que un documento pueda ser compuesto, es decir, que esté formado por más de una pieza periodística, y asignar metadatos a cada una de las posibles partes de un documento NITF: cada objeto (entendido como un fragmento textual con sentido propio), una pieza que contenga más de un objeto, cada una de las piezas, el conjunto de todas las piezas, cada documento periodístico, si hubiera varios en el conjunto del documento NITF y, finalmente, el documento físico NITF completo.

A modo de resumen, la siguiente figura representa una posible estructura física de NITF, mediante sus elementos mínimamente representados, y los contenidos que pueden llegar a tener. En **negrita**, se marcan los elementos estrictamente obligatorios:

<code><nitf></code>	
<code><head></code>	Metadatos de todo el documento
...	
<code></head></code>	
<code><body></code>	
<code><body.head></code>	Metadatos para todo el cuerpo del documento
...	
<code></body.head></code>	
<code><body.content></code>	Pieza informativa (puede haber más de una)
<code><block></code>	Bloque, equivalente a un objeto. Solo es obligatorio si el documento contiene más de un objeto.
(inicio del bloque)	Metadatos del bloque
(contenido central)	Objeto textual
(fin del bloque)	Metadatos del bloque
<code></block></code>	
<code><body.end></code>	
...	
<code></body.end></code>	Metadatos para todo el cuerpo del documento
<code></body.content></code>	
<code></body></code>	
<code></nitf></code>	

Figura 5.22.
Elementos generales de NITF y aplicación de los mismos

Como se observa en el esquema, puede haber más de una pieza informativa en un documento NITF, mediante sendos elementos `body.content`, y cada uno de esos contenidos puede estar a su vez formado por un objeto que contenga elementos estructurales (párrafos, listas, tablas, etc.), o bien, por bloques distintos, cada uno de ellos conteniendo un objeto, con esos mismos elementos.

Un ejemplo de un documento concreto NITF puede ser éste:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE nitf SYSTEM "
http://www.nitf.org/IPTC/NITF/3.2/dtd/nitf-3-2.dtd">
<nitf>
  <body>
    <body.head>
      <headline>
        <h1>El Guggenheim de Bilbao repasa toda la trayectoria
de Kandinsky</h1>
      </headline>
    <istributor>
```

```
<org>LA VANGUARDIA</org>
</distributor>
<dateline>04.04 horas - 05/07/2002</dateline>
</body.head>
<body.content>
  <block>
    <classifier>NOTICIAS</classifier>
    <classifier>CULTURA</classifier>
  </block>
  <block>
    <byline>
      <person>SANTIAGO FDEZ. GUBIEDA</person>
      <location>Bilbao</location>
    </byline>
    <p>El Museo Guggenheim Bilbao recrea la búsqueda de la
    abstracción de Vasily Kandinsky (1864-1944) con una muestra de
    casi medio centenar de obras del pintor ruso y autores coetáneos,
    pertenecientes a la colección Guggenheim. El montaje, que puede
    verse hasta el 12 de enero, bucea en la figura de Kandinsky como
    "artista total", erudito y vanguardista.</p>

    <p>La comisaria de la exposición, Petra Joos, explicó que
    la muestra tiene un doble objetivo: por un lado, mostrar la
    trayectoria artística e intelectual de Kandinsky a través de 41
    cuadros pintados desde sus primeros años en Munich hasta el final
    de sus días en París; y, por otro lado, volver la mirada hacia el
    arte de sus contemporáneos para ver sus influencias y caminos
    paralelos, por medio de 41 obras de 19 autores.</p>

    <p>La primera sala descubre a un Kandinsky asentado a
    principios de siglo en París, donde se dejó seducir por la
    técnica impresionista, aunque pronto surgió el interés por la
    abstracción. La segunda sala está dedicada a la consolidación de
    las formas geométricas: círculos, triángulos y cuadrados. El
    montaje concluye su recorrido en una tercera sala dedicada a las
    influencias surrealistas del pintor, en su última época de
    estancia en París donde entabla relación con Miró y Léger.</p>

  </block>
</body.content>
</body>
</nitf>
```

Figura 5.23.
Ejemplo de documento NITF

5.5.3. Capacidad para la descripción detallada de términos: el texto enriquecido

La tercera peculiaridad de NITF es su doble capacidad de descripción. Como ya hemos visto, por un lado, esta descripción es posible mediante metadatos ubicados en diversas partes de su estructura. Pero además, NITF añade la posibilidad de realizar

la descripción mediante *marcado inserto* de partes muy concretas, palabras o frases, dentro del propio documento.

Esta posibilidad, que se denomina en la terminología de NITF **texto enriquecido**, consiste en el marcado de referencias a protagonistas de la noticia o a personas referidas en la misma, a lugares, acontecimientos, fechas, objetos citados (como obras intelectuales, ya sean artísticas, técnicas, científicas, etc.), direcciones postales o electrónicas, cargos o puestos en una institución, etc. Cada una de estas menciones, a su vez, permite utilizar otros elementos y atributos que restringen aún más la mención concreta. Por ejemplo, es posible indicar que un determinado texto es un nombre propio, y señalar, dentro del mismo, el nombre de pila y los apellidos, como en el siguiente ejemplo:

```
El alcalde de Getafe,  
<person><name.given>Pedro</name.given><name.family>Castro</name.family></person> anunció ayer...
```

Figura 5.24. Ejemplo con el elemento de NITF *person*

En más de una ocasión, los elementos de texto enriquecido pueden combinarse entre sí. Por ejemplo, para indicar dentro del nombre de una persona que está ejerciendo un cargo, podría utilizarse un marcado como el de la figura siguiente:

```
El <person><function>alcalde de Getafe</function>,<br><name.given>Pedro</name.given><name.family>Castro</name.family></person> anunció ayer...
```

Figura 5.25. Ejemplo con el elemento de NITF *function* anidado dentro de *person*

No todos los elementos de texto enriquecido pueden anidarse entre sí. Si además quisiéramos marcar la localización reseñada (en este caso, Getafe), una tercera posibilidad de marcado podría ser ésta:

```
El <function>alcalde</function> de <location>Getafe </location>,<br><person><name.given>Pedro</name.given><br><name.family>Castro</name.family></person> anunció ayer...
```

Figura 5.26. Ejemplo de uso de *function*, *location* y *person*, sin anidamiento

Los elementos de texto enriquecido pueden ir dentro de ciertos elementos estructurales pertenecientes al `body`: titulares, párrafos, listas, citas, menciones de autoría o colaboraciones, y tablas (en este caso, indirectamente, dado que los elementos que componen una tabla pueden incluir a todos los anteriores). Algunos de ellos también pueden ir dentro del elemento `identified-content`, que forma parte del elemento `head`.

Los valores que pueden contener estos elementos, no obstante, no tienen por qué contener texto expresamente escrito en el contenido del documento. El texto enriquecido puede llevar además atributos que permitan asignar códigos o términos concretos tomados de un sistema de clasificación, indicando además cuál es ese sistema de clasificación. Un ejemplo, ficticio, puede ser:

```
<person idsrc="http://www.cargos.es/adminlocal/alcaldes#Madrid" value="28027">Pedro Castro</person> anunció ayer...
```

Figura 5.27.

Ejemplo imaginario de una referencia, desde el elemento `person`, a un código de una lista imaginaria de cargos públicos, que remitiera al alcalde de Getafe (Madrid)

Todos los elementos de un documento NITF pueden llevar, además, un atributo `id`, que permite identificar inequívocamente a cualquier elemento, dado que su valor no puede repetirse en el documento. Este código es especialmente válido para la gestión de los documentos NITF mediante sistemas de gestión de contenido.

5.6. Rasgos documentales de NITF

NITF permite un mayor detalle de descripción respecto a NewsML. Si éste último permitía una estructura cuyo nivel mínimo de detalle era el objeto, NITF permite además hacer descripciones mucho más precisas, tanto de la estructura interna del propio objeto como, a nivel semántico, de cualquier parte del mismo, incluso de una palabra o de una expresión concreta. NITF ofrece la posibilidad, por tanto, de realizar descripciones mediante metadatos, pero además, también cabe la posibilidad de realizar un marcado inserto que permite la contextualización concreta de partes muy precisas. A continuación se explican con mayor detalle ambas posibilidades.

Como se describió en el capítulo 2, un documento NITF dispone en sí de dos partes iniciales: una, formada por el elemento `body`, dedicada a la estructura del contenido, y una inicial, aunque optativa, formada por el elemento `head`, dedicada a la definición de metadatos. No obstante, esta división se ve afectada por otro aspecto ya descrito: la **modularidad**, o capacidad para describir múltiples estructuras textuales. En cada una de ellas, como se muestra a continuación, puede incluirse la correspondiente descripción documental.

5.6.1. Opciones para la descripción documental

Un documento NITF puede tener hasta tres cabeceras distintas: una para todo el documento NITF (mediante el elemento `head`), otra para el elemento `body` (mediante el elemento `body.head`) y una tercera, para cada uno de los bloques que pueden componer el mismo `body` (mediante lo que en la terminología de NITF se agrupa como una *entidad* o conjunto de elementos denominada internamente como `%block.start`, ya descrita como *inicio del bloque* (cf. § 5.5.2.). Esas tres cabeceras pueden contener metadatos, lo cual significa que un documento NITF puede contener hasta tres niveles distintos de metadatos.

Pero, más aún, muchos de los elementos que contienen esos metadatos pueden estar en dos de esas cabeceras, simultáneamente, dependiendo de cada elemento: en `head` y en `body.head`, en `body.head` y en `%block.start`, y en `head` y en `%block.start`.

Esta duplicidad en la posible ubicación de un mismo metadato dentro de un documento NITF tiene sentido en cuanto que dicho documento puede disponer de más de un objeto y de más de una pieza, sin límite alguno en la cantidad, por lo que la descripción genérica de contenidos y de contextos, mediante metadatos a niveles superiores (por ejemplo, para indicar temas genéricos que relacionen a todo un documento) puede ayudar a relacionar todas las piezas que lo componen.

Sin embargo, también cabe el peligro de que los contenidos de dos elementos de metadatos, con el mismo nombre, situados en cabeceras distintas, puedan contener términos que no solo se solapen, sino que se puedan contradecir. ¿Es este aspecto negativo? No necesariamente, dado que una misma pieza, por ejemplo, podría contener dos objetos que se enfrenten entre sí, como dos artículos de opinión que expresen opiniones contrapuestas sobre el mismo tema. En este caso, por lo tanto, sí resulta útil que se contengan descripciones genéricas, por ejemplo, para el conjunto de una pieza (por ejemplo: “aumento de los impuestos”), y descripciones específicas (“mejor reparto de la riqueza” o “abuso de control público”, por citar dos posibles casos) para cada uno de los posibles objetos de texto.

Existe además un cuarto nivel de descripción posible, también basado en la descripción estructural de NITF: el **texto enriquecido**, que permite explicar, mediante marcado inserto dentro del contenido central del documento (es decir, dentro del `body.content`) el contexto relativo a cualquier mención de cualquier detalle concreto en el texto. Esta posibilidad, se aplica a muy pocos elementos, todos ellos insertos dentro del `body...` –excepto en un caso, en el citado elemento `identified-content`, que forma parte del elemento NITF pero de modo adjunto, dentro del elemento `head`, es decir, como un metadato aplicable a un documento NITF en conjunto–. No obstante, estos pocos elementos son los más importantes, desde el punto de vista estructural, puesto que definen el contenido real de la noticia.

5.6.1.1. Descripción inserta en el documento

El uso del texto enriquecido permite que cualquier mención, por pequeña que sea, a un personaje, a un hecho, a un lugar o a cualquier otro factor susceptible de poseer valor periodístico o de cualquier otro tipo, pueda ser marcada y descrita, para cumplir al menos alguna de estas funciones:

- a) **Formalizar cualquier referencia concreta en el texto**, de modo que se pueda reducir la polisemia, y detallar las referencias de sinonimia y otros

tropos (sinécdoque, metonimia) existentes en muchos textos periodísticos. Todos los elementos de texto enriquecido contienen un posible subelemento, `alt-code`, que permite añadir a un texto en lenguaje natural una equivalencia formalizada, es decir, ajustada a un lenguaje documental, y especificar dicho lenguaje del cual se extrae esa referencia.

Si un medio utilizara, por ejemplo, una lista de encabezamientos de materia en la que a cada personaje susceptible de ofrecer interés periodístico se le adjudicara un código de acceso para cada encabezamiento, podrían distinguirse las referencias a, por ejemplo, dos posibles personajes con nombres similares.

```
<p>...<person><alt-code idsrc="personajes.xml" value="#12345"
/>José María Aznar</person> declaró en Bruselas que el
Gobierno...</p>

<p>...<person><alt-code idsrc="personajes.xml" value "#26793"
/>José María Aznar</person> fue multado con 530 euros...</p>
```

Figura 5.28.

Dos ejemplos de menciones a dos personas con el mismo nombre, distinguidas con diferentes referencias formalizadas mediante el elemento `person`.

En el primer caso, la referencia biográfica apunta a José María Aznar, presidente del Gobierno español en 2003. En el segundo, la referencia señala a uno de sus hijos, con el mismo nombre de pila, en una noticia sobre un caso particular. La indicación codificada de que se trata de una persona diferente (mediante un código particular, incluido dentro de la lista contenida en *personajes.xml*) distingue a ambos entre sí.

Esta función no se ve limitada a diferenciar dos menciones similares de un aspecto periodístico (en este caso, dos personajes principales de una noticia con el mismo nombre), sino que también puede ser empleada mediante la especificación de cada una de las menciones con distintos elementos NITF que expliquen semánticamente el papel de cada mención. Por ejemplo:

```
El <person><function>marinero</function>Juan Sebastián  
Elcano</person>...  
  
La reparación del <object.title>Juan Sebastián  
Elcano</object.title>se realizará en el puerto de Cartagena...
```

Figura 5.29.

Dos ejemplos de menciones a dos entidades distintas (un personaje, que ejerce la acción principal, en el primer caso, y un objeto, que recibe dicha acción, en el segundo)

Los dos casos anteriores permiten reducir la polisemia. Pero también sería posible especificar, por ejemplo, dos referencias a la misma entidad, una de ellas mediante una sinécdoque, como en el ejemplo siguiente:

```
El director de cine<person><alt-code idsrc="personajes.xml"  
value="#11818" />Pedro Almodóvar</person>...  
  
El premio otorgado al genial <person><alt-code  
idsrc="personajes.xml" value="#11818" />artista  
manchego</person>...
```

Figura 5.30.

Dos referencias a la misma persona, una de ellas mediante una sinécdoque, especificando la relación entre ambas mediante la asignación de un mismo código

b) Aumentar el nivel de especificidad de una referencia, de modo que se facilite su posterior recuperación documental. Por ejemplo:

```
<person><name.given>Bernat</name.given><name.family>Soria  
</name.family></person>afirmó que...  
  
Un nuevo atentado en <location><city>Toledo</city><region><alt-  
code value="Ohio" /><country><alt-code value="Estados Unidos de  
América" /></country></location>...
```

Figura 5.31.

Dos ejemplos de menciones a dos entidades, un personaje y un lugar, con subelementos que detallan la descripción de ambos

Obsérvese que en el primer caso, la descripción detallada parte del mismo texto incluido en el documento, especificando que se incluye el nombre de pila y el apellido de la persona referida. En el segundo caso, sin embargo, es posible incluir información que no está incluida en el texto para especificar la localización exacta de la ciudad mencionada, de modo que, aunque en el texto no se visualizara esa información, sí fuera susceptible de

ser encontrada con mayor precisión al incluir el documento en un sistema de recuperación documental.

- c) Permitir la localización de información en un **contexto idiomático diferente**. Por ejemplo:

```
...<num units="60.000.000" thousands-ch=".">sesenta
millones</num> de <money unit="US dollars"
date="20030725T120000+01:00">dólares</money> se ha gastado...
```

Figura 5.32.

Referencia a una cantidad y a una moneda en castellano, con atributos formalizados en inglés

En el ejemplo que acaba de referirse, aunque la noticia estuviera en castellano, y la mención a la cantidad, sin normalizar (en este caso, el número está escrito con letras), la noticia podría ser marcada en una base de datos en inglés y de modo formalizado, de modo que un documentalista, por ejemplo, podría buscar ese dato concreto (por ejemplo la venta de un jugador de fútbol de un equipo español, por esa cantidad) en el medio que contuviera esta noticia, incluso sin tener conocimientos de castellano, y ofrecerle esa información a un redactor que hubiera pedido esa noticia en un medio en castellano.

- d) Realizar **menciones a direcciones electrónicas** accesibles desde el documento. Ello puede realizarse mediante el elemento `virtloc`, que describe el texto incluido dentro del elemento como una dirección electrónica, y mediante el elemento `a`, que describe que el texto contenido es un ancla de enlace, es decir, hace referencia a otra parte del mismo documento o a otro documento externo, accesible al realizar una acción (por ejemplo, mediante un clic) sobre dicho texto.

No obstante, conviene aclarar, que, en ninguno de los dos casos, NITF establece mecanismos para que el texto incluido en esos elementos sea interpretado como un ancla real por parte de un navegador, a menos que éste, expresamente, entienda dicho elemento (al igual que hacen los navegadores con el elemento `a` del lenguaje HTML, en el cual se inspira el

equivalente en NITF), situación que actualmente no se ha desarrollado en ningún navegador.

Una posible alternativa sería el uso de XLink (DEROSE; MALER, y ORCHARD, 2001) aplicado a ese elemento concreto, lo cual permitiría además que las capacidades de enlace de documentos NITF fuera mucho mayor que el mero enlace unidireccional definido en HTML, como explica en profundidad (ARELLANO PARDO, 2002), aunque no para el caso concreto de NITF, pero sí para cualquier desarrollo basado en XML.

NITF dispone asimismo de un mecanismo general de texto enriquecido para la descripción de información: el elemento `classifier`, que consiste en un descriptor general sobre cualquier parte del `body.content` de un documento NITF, como elemento hijo del elemento correspondiente. Es decir, `classifier` puede ser un descriptor asignado a una palabra cualquiera del contenido de un objeto, a un elemento de texto enriquecido (como la función de una persona), a un elemento estructural del contenido de una noticia (como un párrafo), a todo un bloque de información incluido en un objeto, o bien, a todo un documento NITF (en este caso, dentro de un elemento `identified-content`, hijo de `docdata`, e incluido a su vez en `head`).

El elemento `classifier` puede contener una descripción en texto libre, o emplear una mención a un término de un lenguaje controlado mediante el citado elemento hijo `alt-code`.

5.6.2. Consideraciones sobre la precisión documental de NITF

El empleo de `classifier` como elemento genérico que puede contener cualquier tipo de información contextual sobre cualquier fragmento de un documento NITF denota la preocupación de los creadores del lenguaje por la inclusión de descripciones añadibles al propio texto. Valga el siguiente ejemplo:

```
Los implicados en <classifier><alt-code idsrc="hechos.xml"
value="#crisis-AsamMad2003">el problema</classifier> han
declarado que...
```

Figura 5.33.
Ejemplo de uso del elemento `classifier`

En el caso anterior, una mención a un hecho noticiable dada por conocida por el autor, al redactar el texto puede ser complementada mediante un término extraído de un lenguaje documental que describa el tema concreto al que se refiera ese texto, que por sí solo puede ser ambiguo, a medio plazo (como la llamada “crisis de la Asamblea de Madrid”, situación causada por la no asistencia a un pleno de la Asamblea de Madrid, en junio de 2003, de dos tránsfugas del PSOE que mantuvieron su escaño posteriormente a pesar de ser expulsados de su partido).

El elemento `classifier`, por otra parte, puede emplearse como elemento de marcado básico por parte de los miembros de la redacción de un medio, a la hora de elaborar un documento, de modo que pudiera servir luego como guía a los documentalistas para indicar que el texto marcado es relevante, a la hora de analizar el documento. Un ejemplo básico de uso podría ser el siguiente:

```
Los implicados en <classifier>la sucesión de Aznar al frente del
PP</classifier> no han defendido aún sus derechos...
```

Figura 5.34.
Ejemplo de uso del elemento `classifier` sin lenguaje documental

No obstante, esta opción, a priori, adolece de cierta falta de precisión documental. Dicha falta se transmite también, en nuestra opinión, con la imposibilidad de combinación entre diversos elementos de texto enriquecido, que podrían ser anidados, o bien, formar parte de un contenedor con el cual amplificar las capacidades de descripción de texto. Por ejemplo, NITF sí permite un marcado como éste:

```
El <person><function>director de la Cadena SER, <person>Antonio
García Ferreras</person></function>
```

Figura 5.35.
Ejemplo de anidamiento de los elementos `person` y `function`

Pero no permite éste otro:

```
El <function>director de <org>la Cadena  
SER</location>...</function>
```

Figura 5.36.
Ejemplo de uso, no válido en NITF, de los elementos *function* y *org*

Y obliga a marcar, si se quieren distinguir ambos elementos, de esta manera:

```
El <function>director</function> de <org>la Cadena Ser</org>...
```

Figura 5.37.
Ejemplo de uso, válido en NITF, de los elementos *function* y *org*

Esta misma situación se da, por ejemplo, a la hora de combinar detalles como un cargo y la institución en la que se desarrolla el mismo (como “presidente de Endesa”), o un acontecimiento relacionado con una institución concreta (en un caso como “regulación de empleo en Telefónica”), entre otras posibles combinaciones.

En la génesis de NITF (ALLEN y MÖHR, 1998), por motivos no suficientemente detallados, se optó en esos casos por la aplicación a todo el conjunto del elemento más específico: si se hacía mención a un elemento concreto en un ámbito mayor, como, por ejemplo, un cargo de una institución, entendiendo que el cargo formaba parte de la institución, se marcaba todo el bloque con el elemento correspondiente al cargo:

```
El <function>presidente de Repsol</function>
```

Figura 5.38.
Ejemplo recomendado de uso en NITF para citar a un cargo de una institución

A nuestro juicio, esta opción, además de inconsistente (dado que sí permite incluir, por ejemplo, *function* dentro de *person*, pero no otras posibilidades), dificulta al menos dos posibles aplicaciones. La primera consiste en el empleo de términos formalizados. Imaginemos que, continuando con el ejemplo anterior, se quisieran introducir términos tomados de un lenguaje documental para ser insertados directamente en el texto dentro de elementos NITF, mediante un programa que pudiera seleccionarlas partiendo de un listado prefijado.

NITF solo permitiría, con la actual configuración, la posibilidad de que ese listado estuviera formado por un lenguaje precoordinado con expresiones como “presidente de Repsol”... con lo cual, si quisiéramos tener la posibilidad de generar una cadena de texto con un cargo y una institución, tendríamos que disponer de un lenguaje documental que contuviera todas las referencias a un cargo y a una institución, al menos, de todos los cargos de instituciones susceptibles de ser utilizados habitualmente: “presidente de Repsol”, pero también, “presidente de Iberia”, o “presidente de IBM”...para que fuera insertadas en sendos elementos NITF.

Por el contrario, si tuviéramos la posibilidad de generar ese elemento NITF anidando contenidos, se tiene la posibilidad de disponer de dos listas jerárquicas, una de ellas con posibles cargos, entre los que se incluya la cadena “presidente”, y otra, con las posibles instituciones más utilizadas, de modo que éstas pudieran usarse indistintamente instituciones al hacer referencia a sus cargos concretos... o a ellas mismas por separado.

De este modo, al yuxtaponer “presidente” con la institución correspondiente, se podría generar de modo semiautomático texto como:

```
El <function>presidente de <org>Repsol</org>...</function>
```

Figura 5.39.
Ejemplo de uso, no válido en NITF, de los elementos *function* y *org*

Si además, se pudieran yuxtaponer más de dos elementos (por ejemplo, cargo, institución y localización), las posibilidades serían mayores, hecho que no es posible en la actualidad.

La segunda aplicación que obvia este enfoque de NITF reside en un caso inverso: la incapacidad de recuperar ciertas referencias a un posible término en el caso de emplear los contenidos de elementos NITF como valores para rellenar posibles campos de una base de datos sobre noticias de prensa.

Supongamos, de modo básico, que en la citada base de datos, se utilizan los contenidos del elemento *function* para introducir información sobre los cargos

implicados en noticias de prensa, y el elemento `org` para indicar qué instituciones aparecen en esas mismas noticias. Si, siguiendo la recomendación de NITF, disponemos de un marcado como el indicado en la figura 5.39, al buscar en el campo de instituciones por el nombre de la misma (Repsol, en este caso), relleno con los contenidos de elementos `org...` no aparecería la noticia en la que figurara ese marcado (salvo que en otro punto de la noticia, apareciera la mención a la institución, por separado, dentro de `org`).

El anidamiento de dos elementos, no obstante (como sí ocurre con `person` y `function`, como se señala en la figura 5.35), puede suponer problemas, dado que junto a los términos empleados, aparecen palabras vacías (en este caso, preposiciones, como “de”, o signos ortográficos, como la coma), con lo cual, en el caso del almacenamiento automático de la base de datos mediante los contenidos de elementos, habría que utilizar herramientas de depuración que eliminaran estas palabras vacías y los signos de puntuación. Sin embargo, un problema mayor sucedería cuando tales palabras no fueran vacías, como en:

```
El <function>alcalde electo de  
<location>Madrid</location>...</function>
```

Figura 5.40.
Ejemplo de inserción de palabras no vacías entre contenido de los elementos `function` y `location`

Una solución alternativa sería la existencia de un elemento contenedor, que permitiera la combinación de dos o más elementos de texto enriquecido.

5.6.3. Hiperflexibilidad estructural de NITF: algunos interrogantes

A todas estas posibilidades se les añade un factor más: ninguna de las citadas estructuras de descripción de cualidades, de contenido o de contexto, ya sean cabeceras de elementos adjuntos (`head`, `body.head`, `&block.start`) o información inserta junto con el texto, son obligatorias. Ello supone que un documento NITF puede optar entre una falta absoluta de descripción o un posible

exceso de la misma. ¿Dónde debe estar el término medio, si es que existe? ¿Qué implicaciones puede tener esta flexibilidad para un posible uso habitual del lenguaje? Y, más aún, si el propio NITF ya ofrece cuatro niveles de descripción, ¿cómo combinarlos con posibles niveles adicionales, como los empleados en NewsML, si se decide integrar ambos lenguajes de marcado?

Éstas y otras preguntas relativas a la utilización de NITF tanto en un centro de documentación como, previa o posteriormente, en los sistemas relacionados con el mismo (redacción, edición, producción, gestión, comercialización) serán desarrolladas en el capítulo 7, a la hora de plantear la inserción de estos lenguajes en el conjunto de un medio periodístico.

5.7. SportsML

El más reciente de los lenguajes IPTC basados en la estructura de los documentos es SportsML: su aprobación final se realizó en mayo de 2003. SportsML está basado en las mismas peculiaridades que NITF y NewsML: modularidad, estructuración y descripción documental de sus contenidos.

Como su nombre indica, SportsML está orientado al uso para contenidos sobre deportes. Su especialidad no se limita al tema elegido: la estructura de este lenguaje está diseñada para contener datos estadísticos, hasta el nivel más detallado, sobre todo tipo de acontecimientos deportivos: participantes, resultados, datos climatológicos, incidencias, modificaciones durante una competición, etc. SportsML está diseñado para recoger todo tipo de hechos sobre deportes, sin importar su duración, ya se trate de un acontecimiento puntual, como un partido de fútbol, o de una competición que se desarrolle durante un largo periodo temporal, como una liga de diez meses de duración.

SportsML describe, en concreto, cinco tipos de acontecimientos deportivos: **encuentros**, o competiciones puntuales entre dos participantes, individuales o colectivos; **torneos**, o competiciones de equipos o jugadores individuales, enfrentados entre sí mediante diversas combinaciones de encuentros; **calendarios**, o

información temporal sobre un conjunto de competiciones relacionadas que forman parte de un torneo; **clasificaciones**, o posiciones de los participantes de un torneo dentro del mismo, según ciertos detalles; **estadísticas**, o datos cuantitativos que comparaban aspectos sobre cualquiera de las anteriores categorías (por ejemplo: número de encuentros de fútbol de una competición concreta en los que un jugador ha marcado algún gol). Finalmente, SportsML permite además incluir lo que denomina **artículos**, o textos periodísticos que comenten alguno de los acontecimientos anteriores. En este último caso, la especificación SportsML recomienda que se use NITF para componer los contenidos de un artículo, aunque puede usarse cualquier otro lenguaje.

5.7.1. Un tema, múltiples variaciones

La especificidad inicial de SportsML supone, en realidad, la asunción de una gran variedad, tanto en la forma como en los contenidos descritos. Dicho de otro modo: no se trata de un único deporte, sino de muchos, cada uno de ellos, con sus distintas reglas y términos. Aunque en todos ellos se producen los acontecimientos citados, las reglas pueden llegar a ser muy diferentes, lo cual determina que los sucesos concretos a describir, y los términos y categorías empleados, también lo puedan ser.

Por otra parte, también es cierto que, al margen de esta diferenciación, pueden señalarse familias de deportes en los que sí se comparten aspectos comunes. Algunos ejemplos: buena parte de los deportes incluyen competiciones mediante el sistema de liga (todos los participantes compiten contra todos entre sí) o por el de copa (se realiza un sorteo que enfrenta a los participantes de dos en dos, y sucesivamente, a los ganadores de cada encuentro entre sí); todos los deportes de pelota se desarrollan mediante encuentros en los que el ganador es aquel que consiga un mayor número de tantos (hacer llegar un esférico hasta un punto defendido por el rival); en las pruebas de atletismo, el viento a favor determina que una marca obtenida sea homologada como un récord, o no; etc.

Esta doble combinación de elementos comunes (los modos de enfrentamiento, la composición de los equipos en deportes colectivos, etc.) y elementos propios (los tipos de aciertos y errores, las sanciones, los lugares en los que se practican, etc.) han llevado a los diseñadores de SportsML a crear un lenguaje modular desde su raíz, creando diferentes estructuras de elementos combinables entre sí, definidas mediante varias DTD.

Todos los documentos SportsML disponen de una estructura básica, denominada DTD núcleo (*Core DTD*), sobre la cual pueden luego añadirse otras estructuras, también en forma de DTD, aplicables a un deporte concreto. En origen, estas DTD están definidas para béisbol, hockey sobre hielo, golf, tenis, fútbol americano, fútbol y baloncesto. Todas ellas han sido desarrolladas por el propio IPTC, y contienen los elementos necesarios para cada deporte.

5.7.2. Estructura básica de un documento SportsML

Un documento SportsML contiene, obligatoriamente, una serie de metadatos iniciales que describen, como mínimo, el deporte al que se refiere el documento y, adicionalmente, referencias concretas sobre el tipo de acontecimiento (encuentro, competición, etc.), sobre qué acontecimiento concreto, qué información se contiene de ese acontecimiento (dónde, cuándo), en qué cantidad, etc. Todo el contenido de estos elementos que contienen metadatos (agrupados en un elemento denominado `sports-metadata`) debe estar basado en lenguajes documentales concretos: ya sean los definidos por IPTC, que ha diseñado todo tipo de listas de autoridades muy específicas para distintos deportes, o bien, usando otros lenguajes documentales concretos, que deben ser referidos mediante un elemento `catalog`: el lenguaje documental al que apunte el contenido de `catalog` debe estar físicamente dentro de un documento NewsML.

Además de metadatos, los documentos en este lenguaje pueden añadir a continuación información concreta mediante marcado inserto en el documento, es decir, datos

mucho más detallados sobre hechos concretos acaecidos en un encuentro, un torneo, etc. Cada uno de los posibles acontecimientos citados incluye un árbol de elementos que, básicamente, pueden contemplarse en la figura siguiente:

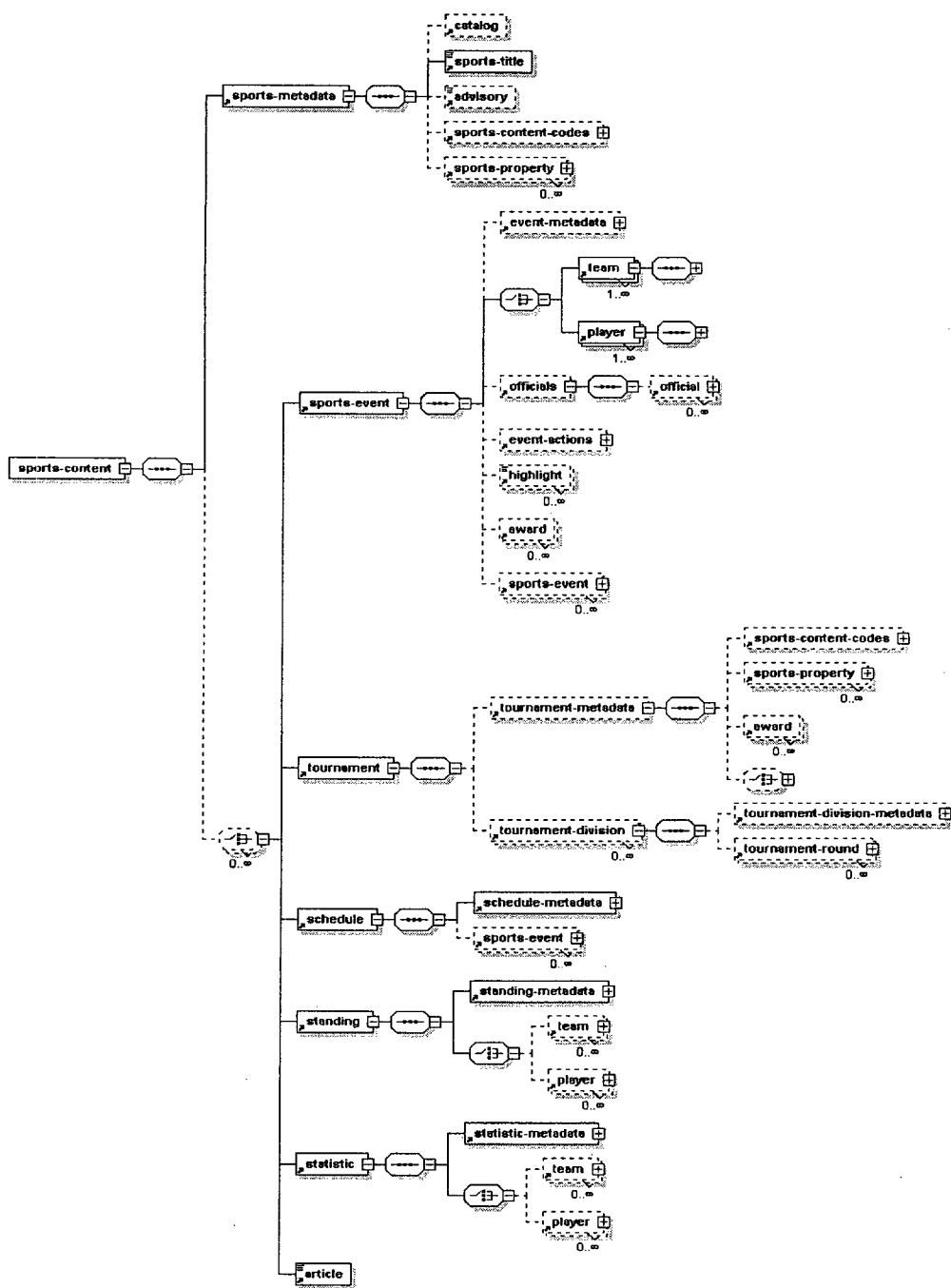


Figura 5.41.
Esquema básico de elementos de SportsML

Por ejemplo, en el caso de que la información sea sobre un encuentro concreto entre dos contendientes, el documento SportsML incluiría un elemento `sports-event`, que podría tener, a su vez, inicialmente una serie de metadatos específicos sobre ese deporte, incluidos datos cuantitativos sobre el encuentro. Además, también pueden ofrecerse datos específicos sobre cada uno de los participantes, ya sean solo dos jugadores o dos equipos completos, incluidos datos de todos sus jugadores. Del mismo modo ocurriría en el caso de que la información fuera sobre una competición, un calendario, una clasificación o una estadística, variando en cada caso el elemento contenedor, utilizando elementos comunes a `sports-event`, pero también añadiendo otros elementos peculiares.

Un mismo documento SportsML puede contener más de un objeto informativo de cualquiera de los tipos mencionados: por ejemplo, pueden insertarse datos de todos los encuentros de una competición, y además, estadísticas determinadas (máximos anotadores, media de puntos por encuentro, etc.). Estos contenidos, así como los citados artículos (textos complementarios comentando un hecho, como la crónica de un partido de fútbol), pueden aparecer en cualquier orden, teniendo en cuenta que, en caso de haber metadatos, ya sean globales o específicos de cada acontecimiento, deberán ir al comienzo del elemento correspondiente.

5.7.3. Uso de lenguajes documentales en SportsML

Dada la gran variedad de los posibles términos empleados, al tratarse de información sobre aspectos prolijamente reglamentados, IPTC ha definido toda una serie de lenguajes documentales controlados, denominados **ficheros de recursos** (Resource Files) (INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL, 2003). Estos lenguajes, que suelen ser listas de autoridades o de encabezamientos de materia, han sido diseñados inicialmente desde una perspectiva anglosajona y, por tanto, en lengua inglesa, desarrollando en particular aquellos deportes mayoritarios en Estados Unidos, con la excepción del fútbol. Los ficheros de recursos contienen, en esencia, tres tipos de términos:

- Acciones propias de deportes concretos. Fundamentalmente, posiciones en el terreno de juego y tipos de penalizaciones.
- Participantes en competiciones. Esta opción es empleada para competiciones por equipos y, sobre todo, cuando se trata de torneos cerrados, como la NBA (baloncesto profesional en Estados Unidos), o son especialmente importantes y sus contenidos serán masivamente utilizados durante un periodo concreto por periodistas de todo el mundo, como los Campeonatos Mundiales de Fútbol.
- Términos susceptibles de ser empleados en futuras variaciones de lenguajes ya existentes o de otros que estén por crearse. El uso de esta opción está restringida a los desarrolladores de SportsML en el seno del IPTC.

Todos estos lenguajes están también desarrollados mediante NewsML. No obstante el uso de dichas listas de autoridades no es normativa, sino solo recomendable por el IPTC, con el fin de normalizar las referencias a un mismo objeto en diferentes medios y, de este modo, mejorar su intercambio entre los mismos.

Además, al igual que en los otros lenguajes de la familia IPTC, todo elemento de un documento SportsML puede incluir un atributo que identifique sin error, mediante un código, al elemento citado. En este caso, el atributo se denomina *i.d.*

5.8. Rasgos documentales de SportsML

SportsML comparte, documentalmente, varias similitudes con NITF y NewsML. La primera de ellas es su **modularidad**, que permite varios niveles de descripción y que se ve desarrollada incluso de un modo físico: a la DTD principal del lenguaje, pueden añadirse además DTD específicas de deportes concretos, como acabamos de señalar.

La principal diferencia de este lenguaje respecto a NITF y a NewsML reside en que un documento SportsML contiene fundamentalmente datos con el fin de facilitar su

tratamiento cuantitativo, mientras que NITF y NewsML están diseñados, aunque no en exclusiva, para la inclusión de documentos, es decir, de contenidos con una menor rigidez estructural. La posible inserción en un documento SportsML de objetos adicionales menos rígidos, como la crónica de un acontecimiento deportivo, deberán ser realizadas en un lenguaje distinto (idealmente debe ser NITF, pero no es obligatorio).

El modelo de descripción empleado para los contenidos, en este caso muy especializados, permite tanto el marcado inserto como el marcado adjunto,; ya sea anexo al conjunto del documento, ya sea anexo a cada uno de los objetos o piezas que lo pueden formar, como al comienzo de cada tipo de hecho deportivo que puede incluido en un documento SportsML: encuentros, torneos, clasificaciones...

En todos estos casos, se permite la correspondiente descripción de prácticamente cada dato que forme parte del documento mediante el empleo de términos extraídos de un lenguaje controlado. Los lenguajes controlados, empleados en documentos SportsML deben estar físicamente en documentos elaborados mediante el formato NewsML.

Esta capacidad de formalización es precisa para el buen funcionamiento del lenguaje, dado que, al estar diseñado con vistas a un eventual procesamiento automático e intercambio de datos deportivos, tanto la estructura como la composición de esos datos requiere una formalización rígida, de modo que el propio documento sea, en lo posible, autoexplicativo.

El uso de SportsML puede llegar a ser muy complejo, dado el número posible de elementos y atributos susceptibles de ser empleados y combinados entre sí. Esta complejidad se debe al gran volumen de información variada que puede desprenderse, en general, de cualquier acontecimiento deportivo, a lo cual se añade el hecho de que cada deporte tiene, en parte o en su totalidad, una terminología muy diferente. Además, como ya se ha señalado, SportsML no describe un solo tipo de hecho deportivo, sino seis. No obstante, gracias a la extrema modularidad de SportsML, también es posible la definición de documentos relativamente sencillos.

En cuanto al conjunto de lenguajes documentales empleados, aunque muy detallados, resultan de la utilización de listas de términos que son, en su mayoría, relativamente estables: al menos, en el caso de las reglas de los deportes y de los hechos deportivos en general. Ello permite que los lenguajes documentales empleados puedan ser susceptibles de ser más fácilmente aceptados por la comunidad periodística internacional, aunque el hecho de que tales términos estén en inglés dificulte su aceptación por medios que no utilicen ese idioma, a menos que consigan utilizar tales términos mediante la correspondiente traducción, ya sea mediante DTD en diferentes lenguas (lo cual iría en contra de la pretensión de uso generalizado del lenguaje en todo el mundo), o bien, mediante una aplicación informática que realice las traducciones desde otras lenguas, de modo transparente para los usuarios del lenguaje.

Referencias

IPTC lanza NewsML. *Técnicas de prensa*, 2000, n° 1, p. 38.

ALLEN, David. *NewsML NewsAgency implementacion guidelines*. 2002. [Consulta: 28/4/2003]. <<http://www.iptc.org/site/NewsML/NewsAgencyGuidelinev1.1c.pdf>>

ALLEN, David y MÖHR, Wiebke. *Considerations for the semantic markup with the NITF*. 1998. [Consulta: 21/6/2003]. <http://www.darmstadt.gmd.de/topas/publications/Allen_Moehr_1998.pdf>

ARELLANO PARDO, M^a del Carmen. *La organización hipertextual del ordenamiento jurídico por medio de tecnologías XML: aplicación a la normativa del IRPF*. Universidad Carlos III de Madrid, 2002.

BANNAN, Karen J. RSS: Lo-Fi Content Syndication. En: *EContent*, 2002. [Consulta: 14/2/2003]. <<http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=837>>

BEGED-DOV, Gabe. *RDF Site Summary (RSS) 1.0*. 2000a. [Consulta: 30/4/2003]. <<http://web.resource.org/rss/1.0/>>

BEGED-DOV, Gary. *RDF Site Summary 1.0 Modules: Dublin Core*. 2000b. [Consulta: 30/4/2003]. <<http://web.resource.org/rss/1.0/modules/dc>>

BURMAN, Linda. *Deadline Duel*. 2002. [Consulta: 26/7/2003]. <<http://www.prismstandard.org/dduel.pdf>>

CLARK, James y DeROSE, Steve. *XML Path Language (XPath) Version 1.0*. 1999. [Consulta: 30/6/2003]. <<http://www.w3.org/TR/xpath>>

COVER, Robin. *News Markup Language (NML)*. 1999. [Consulta: 30/6/2003]. <<http://xml.coverpages.org/nml.html>>

DANIEL, Ron; HANSEN, Derek, y POPE, Cameron. *PRISM: Publishing Requirements for Industry Standard Metadata. Version 1.2 (f). First Public Draft*. 2003a. [Consulta: 30/6/2003]. <<http://www.prismstandard.org/spec1.2f.pdf>>

----- *PRISM: Publishing Requirements for Industry Standard Metadata. Version 1.2 (h)*. 2003b. [Consulta: 29/10/2003]. <<http://www.prismstandard.org/spec1.2h.pdf>>

DeROSE, Steve; MALER, Eve, y ORCHARD, David. *XML Linking Language (XLink) Version 1.0*. 2001. [Consulta: 30/6/2003]. <<http://www.w3.org/TR/xlink/>>

DUBLIN CORE METADATA INITIATIVE. *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. 2002. [Consulta: 28/7/2003]. <<http://es.dublincore.org/documents/dces>>

EL MUNDO.ES. *RSS en elmundo.es*. 2003. [Consulta: 30/7/2003]. <<http://rss.elmundo.es/rss/>>

GETTY RESEARCH INSTITUTE. *Art & Architecture Thesaurus (Research at the Getty)*. 2003. [Consulta: 30/6/2003]. <<http://www.getty.edu/research/tools/vocabulary/aat/>>

HAMMOND, Tony; HANNAY, Timo, y LUND, Ben. *RDF Site Summary 1.0 Modules: PRISM*. 2003. [Consulta: 25/7/2003]. <<http://www2.elsevier.co.uk/~tony/spec/rss/prism.html>>

IANA. *MIME Media Types*. 2002. [Consulta: 28/7/2003]. <<http://www.iana.org/assignments/media-types/>>

IDEALLIANCE. *PRISM Aggregator Document Type Definition (DTD) v. 1.0*. 2003a. [Consulta: 29/10/2003a]. <http://www.prismstandard.org/Pam_1.0/PAMGuide1_0.pdf>

-----, *Industry Milestone: New Standard for Online Content Ready for Prime Time*. 2003b. [Consulta: 29/10/2003b]. <<http://www.prismstandard.org/news/2003/1024.asp>>

International Press Telecommunications Council y Newspaper Association of America. *IPTC-NAA Subject Reference Scheme*. 2003. [Consulta: 30/11/2003]. <<http://www.iptc.org/site/subject-codes/SubjectRefGuidelines.pdf>>

INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (IPTC). *Resource File Documentation*. 2003. [Consulta: 30/6/2003]. <<http://www.sportsml.com/sportsml-documentation/resource-files.html>>

-----, *News Industry Text Format. Document Type Definition - Version 3.1*. 2002a. [Consulta: 30/4/2003a]. <<http://www.nitf.org/site/nitf-documentation/nitf-3-1.dtd>>

-----, *NewsML Document Type Definition Version 1.1*. 2002b. [Consulta: 30/4/2003b]. <<http://www.iptc.org/site/NewsML/DTD/NewsMLv1.1.dtd>>

-----, *IPTC TopicSet-Selector*. 2003a. [Consulta: 13/12/2003a]. <http://www.iptc.org/metadata/mtdta_selform01.php>

-----, *IPTC-TopicSet-List*. 2003b. [Consulta: 13/12/2003b]. <http://www.iptc.org/metadata/mtdta_ts-table01.php>

-----, *Metadata: Subject Reference System & NewsML TopicSets*. [Consulta: 13/12/2003c]. <<http://www.iptc.org/metadata>>

INTERNATIONAL STANDARD ORGANIZATION. *ISO 8601:2000. Data elements and interchange formats - Information interchange - Representation of dates and times*. Ginebra: International Standard Organization, 2000.

-----. *ISO/IEC 4217:2001. Codes for the representations of currencies and funds*. Ginebra: International Standard Organization, 2001.

-----. *ISO/IEC 639-1:2002. Codes for the representations of languages - Part 1: Alpha-2 code*. Ginebra: International Standard Organization, 2002.

LE MEUR, Laurent <lemeur@afp.com>. " Symmetry btw XHTML 2.0 discussions and NewsML/NITF potential evolutions". En: Lista de discusión nitf-standard, 22/5/2003. 2003a [Consulta: 22/5/2003] <nitf-standard@yahoogroups.com>

-----. *NewsML Progress Report*. 2003b. [Consulta: 6/12/2003b]. <<http://groups.yahoo.com/group/newsml/message/1482>>

NACIONES UNIDAS. *Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings*. 2003. [Consulta: 30/6/2003]. <<http://unstats.un.org/unsd/methods/m49/m49regin.htm>>

NASDAQ. *NASDAQ Securities*. 2003. [Consulta: 30/6/2003]. <<http://www.nasdaq.com/asp/symbols.asp?exchange=Q>>

NEWS.COM. *RSS feeds from News.com | CNET News.com*. 2003. [Consulta: 30/7/2003]. <<http://rss.news.com>>

NOTTINGHAM, M. *The Atom Syndication Project (pre-draft)*. 2003. [Consulta: 11/3/2003]. <<http://www.mnot.net/drafts/draft-nottingham-atom-format-00.html>>

REUTERS. *How it all comes together*. 2002. [Consulta: 30/4/2003]. <<http://about.reuters.com/newsml/whynewsml-how.asp>>

SWARTZ, Aaron. *RDF Site Summary 1.0 proposed modules*. 2003. [Consulta: 30/7/2003]. <<http://web.resource.org/rss/1.0/modules/proposed.html>>

U.S. CENSUS BUREAU. *North American Industry Classification System*. 2002. [Consulta: 30/6/2003]. <<http://www.census.gov/epcd/www/naics.html>>

WINER, Dave. *RSS 2.0*. 2002. [Consulta: 30/4/2003]. <<http://backend.userland.com/rss>>