

Análisis prospectivo de las tendencias de desarrollo de los portales periodísticos españoles

Francisco Javier Martínez Méndez

Rosana López Carreño

Grupo de Tecnologías de la Información. Universidad de Murcia (España)

Resumen

El estudio taxonómico de la estructura de los portales periodísticos y el análisis de la frecuencia de aparición de sus componentes —productos informativos, productos documentales y servicios de valor añadido— han fijado el actual estado de la cuestión sobre su desarrollo. La minería de datos es una tecnología que permite identificar tendencias y patrones seguidos en ese desarrollo a partir de la base de conocimientos aportada por los trabajos anteriormente llevados a cabo, permitiendo recuperar información suplementaria a la ya explicitada. Aplicando un análisis multivariante a la matriz de portales periodísticos y elementos componentes, se obtiene una serie de agrupamientos que permitirán determinar qué portales marcan la tendencia a seguir, al mismo tiempo que se aporta información adicional sobre sus grupos de comunicación y sobre el medio origen del que proceden —prensa, radio, televisión o Internet—. El positivo resultado de esta investigación verifica su viabilidad, vaticinando su posible adaptación a otros tipos de portales desarrollados en la web.

Palabras clave: Minería de datos. Recuperación de información. Análisis multivariante. Técnicas de *clustering*. Portales periodísticos. Productos informativos. Productos documentales. Servicios de valor añadido.

Abstract

The taxonomical study of the structure of journalistic web portals and of the frequency analysis of their elements —informative products, documentary products and added value services— have centered the reviews about their development. Data mining is a technology that allows the identification of trends and patterns followed in this development. This begins from the original knowledge base conformed by the group of works previously developed, and allows the recovery of additional information. Applying a multivariate analysis to the matrix formed by the Journalistic portals and their elements, we will obtain several series of clusters that will allow to determine which journalistic portals establish the trend to follow, at

the same time that gives additional information about their communication groups and about the original mass media related to the portal —press, radio, television or Internet—. The positive results extracted from this research verify its viability, predicting possible applications of this methodology to other types of web portals.

Keywords: Information research. Data mining. Information retrieval. Multivariate analysis. Clustering techniques. Journalistic portals. Informative products. Documentary products. Added value services.

1. Introducción

El estudio de los portales periodísticos españoles y el análisis de su composición a partir de la taxonomía de sus elementos componentes constituyen una novedosa e interesante línea de investigación que permite una adecuada extracción de conocimiento sobre el grado de desarrollo de estos sistemas de información. La web permite que el usuario adquiera información de las dos maneras: por medio de una ecuación de búsqueda y de la navegación a través de hipervínculos (Baeza-Yates, 1999). Ambas modalidades de recuperación de información suelen realizarse desde una página de inicio definida por defecto en nuestro navegador. Muchas veces, esta página es un portal. *¿Qué se entiende exactamente por portal?* Su definición más común es la de concentración de productos y servicios en una misma página, desde la cual se permite el acceso a las distintas secciones de una web. A diferencia de una página tradicional que solo ofrece información particular de una empresa o institución y sobre la que eventualmente se hacen modificaciones a su contenido, los portales, además de proveer contenidos propios, son sitios dinámicos que facilitan transacciones e interactividad con el usuario, por medio de una oferta de servicios tan variada como el comercio electrónico, el correo electrónico, la creación de comunidades de usuarios con intereses afines, juegos, etc. En definitiva, los portales son sitios web que actúan como punto de entrada a Internet, concentrando productos y servicios y organizando la localización de otros sitios, ayudando con ello a los usuarios.

El primero y más famoso portal desarrollado desde casi el comienzo de la web es *Yahoo!*, al que siguieron otros muchos sitios web generalistas o especializados (estos últimos también conocidos como “vortales” o “portales verticales”). La segunda fase de popularización de los portales coincidió con el desarrollo de los sitios de los principales operadores telefónicos o ISP (*AOL, MSN, Vodafone, Tiscali, Terra*, etc.), de los portales de empresas informáticas (*Microsoft, Sun Microsystems, Hewlett Packard*, etc.), y de sitios web dedicados al comercio electrónico o a las subastas (*eBay, Amazon*, etc.). Algunos de estos sitios aún son de los más visitados de la web (1), ya que suministran a sus usuarios una amplia serie de noticias y servicios de muy diverso tipo. Al mismo tiempo que las empresas del sector tecnológico han consolidado su presencia a través de los portales, hemos asistido a la

implantación progresiva de los portales periodísticos, sitios que han evolucionado desde unas sencillas ediciones digitales de parte del diario impreso hasta erigirse en verdaderos portales que agrupan tanto la información típica del medio de comunicación social como una variada y cada vez mayor oferta de servicios. Esta evolución ha ido pareja al aumento de su índice de aceptación entre los usuarios de la web. En el caso español, según los datos del Estudio General de Medios de abril-mayo de 2005 (2), Internet ha alcanzado dentro de los medios de comunicación un índice de penetración del 18,8% (en 1997 era del 0,9%), aunque aún queda algo lejos de los índices de la prensa escrita, pero en continuo crecimiento, lo que puede dar una idea de la importancia que van adquiriendo estos portales ante los gestores de las empresas de comunicación.

Si bien existe una considerable bibliografía dedicada en el campo de la comunicación a cómo adaptar las técnicas tradicionales del periodismo al nuevo medio (McAdams, 1995; Harper, 1998; Kenney, Gorelik y Mwangi, 2000; Flores y Arruti, 2001; Deuze, 2001; Díaz, 2001; Peñafiel y López, 2002; Medialabs, 2002), no es tan profusa la aportación procedente desde el campo de la documentación sobre este fenómeno. Destaca la realizada por la profesora Fuentes i Pujol y sus colaboradores de la Universidad de Barcelona (Fuentes, Jiménez y González, 2000; Jiménez, Fuentes y González, 2002), y la elaborada por nuestro grupo de investigación alrededor de la tesis doctoral *Análisis de los portales periodísticos españoles. Taxonomía de sus elementos componentes* (López Carreño, 2003), trabajo acompañado de desarrollos paralelos y posteriores (López Carreño y Martínez Méndez, 2002, 2003; Martínez Rodríguez, 2003; Guerrero, 2003; Poblaciones, 2003) y (López Carreño, 2004). Esta relativa poca producción científica y su alto grado de concentración son circunstancias extrapolables al área de los estudios llevados a cabo hasta ahora sobre los portales en general, que quizá han suscitado más interés entre los especialistas de los negocios, marketing y publicidad que entre los documentalistas, a pesar de la fuerte presencia del componente informativo en un portal (en realidad, las aportaciones realizadas en España desde el campo de la documentación se han centrado más en la realización de trabajos de referencia que en el análisis de la estructura y elementos de los portales).

En este tipo de portales pueden identificarse claramente tres pilares sobre los que se asienta su estructura: los productos informativos, los productos documentales y los servicios de valor añadido (López Carreño, 2003). Los resultados de trabajos anteriores permiten afirmar que estos portales siguen una evolución algo lenta pero constante; que los productos informativos constituyen la esencia de estos portales y que son predominantes frente a los documentales. El portal más desarrollado en cuanto a su oferta de productos y servicios es *El País*. No se han detectado grandes diferencias entre los portales de alcance nacional y los de alcance regional y curiosamente no existen grandes diferencias entre los portales de origen prensa

y los de origen radio y televisión. El análisis y sistematización de esta tipología de componentes representa la base necesaria para la extracción posterior de estas conclusiones y también puede servir como punto de referencia para el desarrollo de proyectos de la misma naturaleza en otro tipo de portales, quizá algo más generalistas en su alcance y no tan ligados a los medios de comunicación, pero que comparten gran parte de elementos comunes.

Sin restar mérito a la determinación del estado actual de desarrollo de los portales periodísticos, el interés desde el campo de la Documentación no acaba ahí, sino que contempla otros aspectos a considerar, para lo que será necesario realizar nuevos trabajos de investigación. Dos son los principales puntos a analizar en estos trabajos, la *usabilidad* de los portales periodísticos españoles y la identificación de posibles *tendencias de desarrollo*. La usabilidad de los sitios web, es decir, “la utilidad de un sistema, en tanto que es un medio para conseguir un objetivo” (Nielsen, 1993), es uno de los aspectos más estudiados hoy en día en nuestro campo. Lógicamente, los portales periodísticos no pueden escapar a esta tendencia y no deja de sorprender la escasísima producción realizada hasta ahora, resaltando las aportaciones de Donaghy (2002) y García Gómez (2004). Por tanto se trata de un área de estudio aún inexplorada. Actualmente nuestro interés se centra en conferir una mayor envergadura al citado trabajo de García Gómez, donde se analizan las páginas principales de los portales periodísticos de los diarios de prensa más importantes de España, junto con algunos portales de diarios totalmente digitales (que no poseen versión impresa). Los resultados obtenidos hasta ahora adelantan que las páginas principales de los portales de los diarios creados directamente para Internet poseen una mayor usabilidad que las pertenecientes a los tradicionales diarios impresos, sistemas algo más encorsetados por sus guías de estilo y de imagen corporativas, que suelen trasladar al nuevo entorno de la web (estas conclusiones vienen a resaltar una vez más las características intrínsecas de la web en todo lo relacionado con la interactividad y la presencia de una cierta “complicidad” con el “usuario digital”, aspecto que no parece haber desaparecido del todo).

La identificación de posibles tendencias o modelos para el desarrollo de los portales periodísticos constituye actualmente el otro de los principales puntos de interés de nuestra actividad investigadora en este campo. Estamos plenamente convencidos de que la aplicación de diversas técnicas de *análisis de datos de naturaleza multivariante* (técnicas introducidas varias veces en el campo de la recuperación de la información (Jardine y Rijsbergen, 1971; Rijsbergen, 1999; Tombros, Villa y Rijsbergen, 2002), nos permitirá extraer una información suplementaria a los datos generados por las revisiones de la presencia de los distintos productos y servicios en los portales, formalizando así un procedimiento de establecimiento de tendencias de desarrollo de los portales periodísticos con base en sus componentes, transportable fácilmente a cualquier otro entorno de la web, y aportando una serie

de resultados adicionales a los obtenidos previamente, lo que conlleva un alto valor añadido de gran utilidad.

2. Naturaleza, objetivos y etapas del estudio

Si bien la *minería de datos* es una de las técnicas de investigación más empleadas en la actualidad, tal como indica Sánchez Ovcharóv, “es una nueva tecnología punta para uno de los procesos más viejos de los esfuerzos humanos: el reconocimiento de patrones” (Sánchez Ovcharóv, 2001). Thearling la define como una tecnología que lleva a cabo “la extracción automática de información predictiva que se encuentra oculta en grandes bases de datos” (Thearling, 2001). Este autor pone mucho énfasis en que se la considere como una tecnología en sí misma, no como una solución de negocios. En la actualidad, muchos sistemas de información se encuentran inundados con datos generados diariamente por las transacciones con sus usuarios (las visitas a un sitio web, los códigos de barras de supermercados, las operaciones de tarjetas de crédito, las llamadas telefónicas, las recetas de medicamentos, etc.), y sus gestores afrontan el reto del reconocimiento de patrones, perfiles y tendencias presentes y significativos a través del análisis de los datos, para proponer nuevas ideas sobre el diseño del sitio web, sobre una determinada línea de productos, sobre un nuevo tipo de medios de pago o datos ciertos sobre un gasto excesivo y quizá injustificado en algún tipo de medicamento, con el propósito de que los responsables ejecutivos dispongan de más información a la hora de la toma de decisiones. Para muchos autores la gestión de este tipo de información adicional posee un valor superior al de la información almacenada en las bases de datos, y la clasifican como *gestión del conocimiento*; de hecho, la minería de datos es una de las tecnologías más empleadas en este campo. Thearling identifica dos grandes áreas de aplicación de la minería de datos: (I) la predicción automática de tendencias y comportamientos y (II) la identificación automática de patrones de comportamiento desconocidos hasta ahora.

En nuestro caso disponemos de una abundante información sobre el estado de desarrollo de los principales portales periodísticos españoles y pretendemos que la misma devengue un nuevo producto: el conocimiento de las tendencias de desarrollo de estos sitios web. Así, nuestro trabajo tiene como objetivo principal demostrar la hipótesis de que a partir de la observación de la presencia de los distintos elementos componentes de un portal periodístico es posible establecer tendencias en su desarrollo. Este análisis de tendencias se realizará bajo un doble prisma. En primer lugar se intentará advertir la existencia de posibles agrupamientos de portales periodísticos de características muy comunes en su desarrollo (independientemente de su medio origen o de su temática), y, en segundo lugar, se intentará analizar la existencia de concentraciones de productos y servicios agrupadas alrededor de los portales más destacados y que marcan esta tendencia, con el fin

de precisar qué tipo de productos o servicios poseen un mayor valor de discriminación. Para la identificación de estos agrupamientos suelen emplearse técnicas del análisis de *clusters*, consistentes en la identificación de objetos similares dentro de un espacio multidimensional. Se trata de una técnica englobada dentro del campo del análisis multivariante o “conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado. Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir” (Gondar, 2002). Estos métodos se desarrollaron inicialmente para su uso en las ciencias naturales y se han utilizado posteriormente en un amplio número de campos, incluyendo entre ellos el de la recuperación de información y el de la gestión del conocimiento. Esta multidisciplinariedad les viene conferida porque la práctica de clasificar objetos, otorgando la percepción de similitudes, es la base para muchas ciencias. La organización de datos en agrupaciones sensatas es una de las maneras más simples y fundamentales de entender y aprender; por tanto, las técnicas de *cluster* se constituyen en unos poderosos aliados cuando se trata de aplicar un proceso de minería de datos para ampliar el nivel de conocimiento sobre un tema.

Gondar Nores establece tres grandes grupos para clasificar a las técnicas de análisis multivariante (Gondar, 2002).

- a) *Métodos de dependencia*. Se supone que las variables analizadas están divididas en dos grupos: las dependientes y las independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.
- b) *Métodos de interdependencia*. Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué.
- c) *Métodos estructurales*. Suponen que las variables están divididas en dos grupos: el de las dependientes y el de las independientes. El objetivo de estos métodos es analizar no solo cómo las variables independientes afectan a las dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

En nuestro caso resulta imposible establecer a priori qué variables son dependientes y cuáles son independientes. Si bien es lógico suponer que algunos servicios de valor añadido dependen de la existencia de algún producto para poder desarrollarse (especialmente aquellos servicios englobados dentro de la categoría de *acceso a la información*), el hecho de que esos productos existan no implica ne-

cesariamente el posterior desarrollo del servicio. Además, la información de la que disponemos habla de la presencia (o ausencia) de un componente en un portal determinado y nuestro objetivo es establecer qué portales están relacionados, cómo lo están y por qué, por lo que resulta preciso aplicar *métodos interdependientes* para extraer esa información. Dentro de este grupo de métodos está el análisis de *clusters*, cuyo objetivo es “clasificar una muestra de entidades (individuos o variables) en un número pequeño de grupos de forma que las observaciones pertenecientes a un grupo sean muy similares entre sí y muy disimilares del resto” (Gondar, 2002).

La etapa inicial de un proyecto de esta naturaleza es la recogida de datos. Para ello disponemos de una serie de trabajos previamente desarrollados, estudios que aportan los datos que necesitamos (López Carreño, 2002, 2003, 2004; Martínez Rodríguez, 2003). Estos trabajos proporcionan información precisa sobre el desarrollo de productos y servicios en los portales periodísticos españoles correspondientes a los principales diarios de prensa regional y nacional, radio y televisión. A este conjunto de portales le añadimos los dos portales periodísticos realizados directamente en Internet más destacados en cuanto a su usabilidad (García Gómez, 2004). En definitiva, la muestra a estudiar abarca a 32 portales periodísticos españoles, lo que equivale a un tercio de los desarrollados, y en ella figuran sin duda alguna los principales del sector (3).

ABC	ABC	El País	EPS	Onda Cero	CERO
Hoy	HOY	El Mundo	EMU	Cadena 40	40
Vocento	VOC	Expansión Directo	EXP	Cadena 100	100
La Vanguardia	VAN	Marca	MAR	Televisión Española	TVE
Mundo Deportivo	MUN	El Periódico	EPE	CNN+	CNN+
Información	INF	Sport	SPO	Antena 3	A-3
Faro de Vigo	FAR	Panorama	PAN	Telecinco	T5
Nueva España	NES	COPE	COPE	Canal +	PLUS
Levante Digital	LEV	Radio Nacional	RNE	Vilaweb	VIL
5 Días	5DI	Radio Cable	RCA	Diario Directo	DIR
As	AS	Cadena SER	SER		

Tabla I. Muestra de los 32 portales periodísticos españoles objeto de estudio: nombre y acrónimo

Una vez establecida la muestra objeto de estudio, se han de trasladar los datos recopilados a una matriz de portales y componentes, que permitirá desarrollar la segunda etapa de nuestra investigación, la identificación de los posibles agrupamientos de los portales en torno a unos modelos comunes de desarrollo. Esta fase

de estudio puede además proporcionarnos información sobre la evolución de los portales de los distintos grupos de comunicación y sobre el nivel de desarrollo según sea su medio origen (prensa, radio, televisión o Internet). Los resultados de esta fase de nuestra investigación pueden ser contrastados con los aportados en los trabajos anteriores, con el fin de validar nuestro método de observación.

2.1. Estructura de los portales periodísticos

En la estructura de los portales periodísticos, se pueden distinguir tres tipos de componentes fundamentales, a través de los cuales “el contenido es canalizado por medio de los productos informativos, productos documentales y servicios de valor añadido” (López Carreño, 2003). El producto informativo es aquel “producto vinculado directamente a la unidad básica, la noticia”, y el producto documental es “continente de información tratada y organizada, no necesariamente vinculada a la noticia, cuya finalidad es satisfacer una necesidad informativa definida a priori”. El producto informativo constituye un refuerzo de la noticia mientras que el producto documental refuerza otro tipo de informaciones, aunque en muchas ocasiones estos conceptos se entremezclan y es complicada su delimitación, existiendo una frontera difusa entre ellos. También es a veces difícil distinguir claramente entre producto y servicio, especialmente cuando proveen acceso a la información. La tabla II recoge los productos informativos y los productos documentales que van a formar parte de nuestra guía de observación para la realización del estudio (4).

Categoría	Tipo de producto
Producto informativo	pi1 -Noticias; pi2 -Especiales; pi3 -Suplementos; pi4 -Edición impresa; pi5 -Entrevistas; pi6 -Programación TV y radio; pi7 -Tiempo; pi8 -Loterías
Producto documental	pd1 -FAQ; pd2 -Agenda; pd3 -Guías de recursos; pd4 -Guías turísticas; pd5 -Callejeros; pd6 -Asesoría; pd7 -Empleo; pd8 -Resúmenes de noticias

Tabla II. Productos informativos y documentales más habituales de los portales periodísticos. [Fuente: López Carreño, 2003]

Son varios los autores que intentan establecer categorías de servicios de valor añadido. Para Valenzuela y Sallent son aquellos “que se elaboran o prestan con un alto grado de correspondencia con los atributos de las necesidades informativas” (Valenzuela y Sallent, 2000) ya que, en la práctica, el suministro de servicios con valor añadido requiere mayor elaboración. Para Fuentes i Pujol son los elementos “que a las versiones en papel les sea técnica y físicamente imposible ofrecer” (Fuentes, Jiménez y González, 2000). La mayoría de estos portales ofrecen una amplia gama de servicios adicionales a los contenidos puramente periodísticos con el objetivo de atraer y mantener el tráfico hacia su web y complementar su oferta. Así,

“el valor añadido representa, por tanto, un registro opcional, de pago o gratuito, a partir del cual el usuario requiere una atención personal por esa aportación económica o información suministrada” (Palazón Mesguer, 2001).

Aun así, debemos reiterar que a veces no resulta fácil establecer cuándo un componente es un producto y cuándo es un servicio de valor añadido. López Carreño (2003) recuerda que “los servicios de valor añadido (SVA), se identifican como aquellos servicios que resaltan sobre la oferta común de los portales, en general, por implicar un valor extra y que en este caso se desarrollan bajo las herramientas que Internet posibilita hoy, así lo que hoy se considera SVA probablemente mañana será un servicio común, es decir, dependiendo de la condición espacio-tiempo en la que los SVA emergen”. A pesar de estas dificultades, en el desarrollo de su tesis doctoral, la autora es capaz de determinar hasta 45 tipos distintos de servicios de valor añadido agrupados en ocho categorías, que pueden observarse en la tabla III (5).

Categoría	Tipo de servicio
Aceso a la información	sv11 -Ayuda; sv12 -Barra de navegación/Página de inicio; sv13 -Buscador noticias; sv14 -Buscador web; sv15 -Hemeroteca; sv16 -Herramientas; sv17 -Mapa web; sv18 -Portales verticales/Canales; sv19 -Últimas noticias
Interactivos	sv20 -Chat; sv21 -Foros; sv22 -Webmail
Participación	sv30 -Cartas al director; sv31 -Concursos; sv32 -Encuestas; sv33 -Envío de relatos; sv34 -MailNew a un amigo; sv35 -Página abierta; sv36 -Postales; sv37 -Sugerencias
Multimedia	sv41 -Fotografías; sv42 -Gráficos interactivos; sv43 -Radio digital; sv44 -Vídeos; sv45 -TV digital
Personalizados	sv51 -Alertas; sv52 -MailNews; sv53 -Mi diario; sv54 -Mi portada; sv55 -Recordatorios (eventos); sv56 -WAPnews; sv57 -Mascotas virtuales
Entretenimiento	sv61 -Horóscopos; sv62 -Juegos/Pasatiempos; sv63 -Webcam
Comerciales	sv71 -Tienda/s; sv72 -Viajes; sv73 -Clasificados; sv74 -Venta de contenidos; sv75 -Banca online
Complementarios	sv81 -Diccionario; sv82 -Enlace a otros medios; sv83 -PDAnews; sv84 -Traductor

Tabla III. Categorías de servicios de valor añadido más habituales de los portales periodísticos. [Fuente: López Carreño, 2003]

2.2. Desarrollo del estudio

De la observación detenida y rigurosa de la presencia de todos estos componentes en los 32 portales periodísticos miembros de la muestra, se obtienen los datos que permiten la construcción de la matriz de portales y de sus elementos componentes (se recoge una sección en la tabla IV). Como también va a ser objeto de es-

tudio de nuestra investigación el desarrollo de los grupos de comunicación, se ha obtenido la matriz de grupos y elementos componentes (una sección de la misma puede verse en la tabla V).

	pi1	pi2	pi3	pi4	pi8	pd1	pd2	pd3	pd8	sv11	sv12	sv19	sv21	sv22			
ABC	1	1	1	1	..	1	0	0	0	..	0	0	0	..	1	1	..
HOY	1	0	1	0	..	1	0	1	1	..	0	0	0	..	1	0	..
VOC	1	1	1	1	..	1	0	1	1	..	0	0	1	..	1	1	..
VAN	1	1	1	0	..	1	0	1	0	..	0	0	1	..	1	1	..
MUN	1	1	1	1	..	0	0	0	1	..	0	1	0	..	0	1	..
INF	1	1	1	1	..	1	0	1	0	..	0	0	0	..	0	0	..
FAR	1	0	1	1	..	1	0	1	0	..	0	0	0	..	0	1	..
NES	1	1	1	1	..	1	0	1	0	..	0	0	0	..	0	1	..
LEV	1	1	0	1	..	1	0	0	1	..	0	1	1	..	0	0	..
5DI	1	1	1	1	..	1	1	0	1	..	0	0	1	..	0	0	..
AS	1	1	1	1	..	1	0	0	1	..	1	1	1	..	1	1	..
EPS	1	1	1	1	..	1	1	1	1	..	0	0	1	..	1	1	..
EMU	1	1	1	1	..	1	1	1	1	..	1	1	1	..	0	1	..
EXP	1	1	1	1	..	0	0	0	1	..	1	0	1	..	1	1	..
MAR	1	1	1	1	..	1	0	1	1	..	0	0	0	..	1	1	..
EPE	1	1	1	0	..	1	0	0	0	..	0	0	0	..	1	0	..
SPO	1	0	0	1	..	1	0	0	1	..	0	0	1	..	1	1	..
PAN	1	1	0	0	..	1	0	1	0	..	0	0	1	..	1	1	..
COPE	1	1	0	0	..	0	0	0	1	..	0	0	0	..	1	0	..
RNE	1	0	0	0	..	0	0	0	1	..	0	0	0	..	0	0	..
RCA	1	1	1	0	..	0	0	0	0	..	1	1	0	..	0	0	..
SER	1	1	0	0	..	0	0	0	1	..	1	0	0	..	0	1	..
CERO	1	0	1	0	..	0	1	0	1	..	1	1	0	..	0	0	..
40	1	1	1	0	..	0	1	1	1	..	0	1	0	..	0	1	..
100	1	1	0	0	..	0	0	0	1	..	0	0	0	..	0	0	..
TVE	1	0	0	0	..	0	0	0	1	..	1	0	0	..	0	0	..
CNN+	1	0	1	0	..	0	0	0	0	..	0	1	0	..	0	1	..
A3	1	0	0	0	..	0	0	0	0	..	0	0	0	..	1	1	..
T5	1	0	0	0	..	0	0	0	0	..	0	0	0	..	0	1	..
PLUS	1	1	0	0	..	0	0	0	1	..	0	0	0	..	0	1	..
VIL	1	1	1	0	..	0	0	0	1	..	1	0	0	..	1	1	..
DIR	1	1	0	0	..	1	0	1	0	..	1	0	0	..	1	0	..

Tabla IV. Sección de la matriz de portales y elementos componentes.

	pi1	pi2	pi3	pi4	pi5	pi6	pi7	pi8	pd1	pd2	pd3	pd4	pd5	pd6	pd7	pd8	sv11	sv12	sv13	sv14	sv15	sv16	sv17	sv18	sv19	sv21	
VOC	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	
VAN	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	
IBE	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	
PRI	1	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	
REC	1	1	1	1	1	0	0	1	0	1	1	0	0	1	0	1	0	1	1	0	1	1	0	1	1	1	
ZET	1	1	1	1	1	1	0	1	0	0	1	0	1	1	0	0	0	1	1	0	1	1	0	1	1	1	
COPE	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
TVE	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	
A-3	1	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	1	1	

Tabla V. Sección de la matriz de grupos de comunicación y elementos componentes

A modo de síntesis, los trabajos anteriores proporcionaron las siguientes conclusiones (López Carreño, 2003, 2004):

1. Los productos informativos prevalecen en su implantación sobre los documentales, seguramente por el alto grado de vinculación existente de forma tradicional entre estos productos y el medio donde se desarrollan (portales de medios de comunicación).
2. El grado de desarrollo de los servicios de valor añadido es muy heterogéneo.
3. Los portales correspondientes a la prensa tienen un mayor nivel de progreso que los de origen radio o televisión.
4. No se aprecian diferencias significativas entre los portales, independientemente de su origen. De hecho, los portales de radio y televisión comparten la mayoría de los elementos.

Igualmente, en López Carreño (2003) se introduce un “modelo empírico de desarrollo de portal periodístico”, basado en tres niveles de desarrollo: inicial, intermedio y avanzado, según el número de productos y servicios presentes en su oferta. Siguiendo este modelo, solo el portal del diario *El País* podría considerarse “avanzado”, seguido a cierta distancia de algunos portales correspondientes a diarios de prensa. Los portales de radio y televisión en su mayoría se encuentran en un nivel inicial y los portales de prensa se distribuyen entre inicial y moderado. Hay un aspecto en este criterio de clasificación que puede ser objeto de crítica, porque situándonos en el contexto de la profesión periodística cabría la duda razonable sobre lo acertado de asignar igual grado de importancia a los servicios de valor añadido que a los productos informativos, ya que no podemos olvidar que estos son la esencia del portal periodístico (son los elementos más íntimamente

relacionados con la noticia). La respuesta a esta pregunta no es tan simple, porque si nuestra idea es analizar el grado de desarrollo de unos portales y estos servicios se encuentran presentes en su oferta, no sería correcto ni procedente desdeñarlos del todo por su menor vinculación con el tradicional contexto periodístico.

La tabla VI muestra una clasificación de los portales basada exclusivamente en la frecuencia de aparición de sus elementos componentes. De los 60 elementos posibles, el portal del diario *El País* alcanza la cifra de 46, seguido del portal *Vo-cento* con 40 y el portal del *El Mundo* con 39. Solamente cinco portales más superan la cantidad de 30 elementos (la mitad de los posibles), lo que equivale a decir que solo una cuarta parte de la muestra analizada (8 portales) supera la mitad de elementos componentes. Este panorama poco alentador podría cambiar si se reconsiderara de alguna manera la importancia relativa de los servicios de valor añadido frente a los productos informativos (los más frecuentes en los portales periodísticos) y frente a los productos documentales (que en algunos grupos también están muy insertados aunque en otros no tanto). No obstante, el plantear un sistema de asignación de pesos que favoreciera a los productos en detrimento de los servicios podría implicar la inherente inserción de matices subjetivos en nuestro estudio, algo no excesivamente recomendable.

EPS	46	VAN	33	SDI	29	HOY	21	MUN	19	DIR	18	CERO	14	COPE	11
VOC	40	MAR	33	40	27	EPE	21	VIL	19	INF	17	A-3	14	100	7
EMU	39	ABC	32	EXP	24	SER	21	FAR	18	RCA	15	PLUS	14	TVE	7
AS	36	PAN	32	NES	22	LEV	20	SPO	18	CNN+	15	T5	12	RNE	6

Tabla VI. Frecuencia de aparición de los elementos componentes en los portales periodísticos

Las muestras de datos de las que disponemos son relativamente pequeñas, especialmente en lo que se refiere al número de objetos a analizar, que totalizan 32. Este número se reduce considerablemente, a un total de 9 objetos, en la muestra de grupos de comunicación. Las técnicas de *clustering* jerárquico son las más adecuadas para trabajar con muestras de este tamaño. En el caso de muestras más numerosas es recomendable el procedimiento denominado *Cluster de K-medias* (Visauta y Martori, 2003). El *cluster* jerárquico parte de la construcción de una matriz de distancias entre los distintos vectores de la matriz de portales y elementos (los vectores se corresponden con los portales), procediendo a realizar agrupamientos siguiendo un criterio aglomerativo por medio del cual se van “agrupando los objetos o variables en clusters cada vez más grandes hasta que todos ellos forman parte de uno único, de modo que cuando comienza el análisis tendremos tantos clusters como sujetos y al final uno único” (Visauta y Martori, 2003).

El siguiente paso de nuestro procedimiento de análisis de clusters jerárquicos consiste en la determinación de la matriz de distancias entre los distintos portales periodísticos. Todos los cálculos se realizarán con el paquete SPSS. Dentro de este tipo de análisis se puede elegir entre una amplia serie de medidas que se diferencian en función del tipo de datos para el que han sido creadas (cuantitativos, categóricos o dicotómicos), o en función del tipo de distancia evaluada (similitud o disimilitud). Las medidas de similitud evalúan el grado de parecido o proximidad entre dos elementos, enfatizando en el grado de diferencia o lejanía entre dos elementos. Los valores más altos de disimilitud indican mayor diferencia. Son estas medidas las que se emplean comúnmente para construir la matriz de distancias, siendo la más utilizada *Euclídea al Cuadrado* (equivalente al segmento lineal que une dos elementos y que se calcula por medio del sumatorio de los cuadrados de la resta de los valores *i*-ésimos de cada vector dentro del espacio multidimensional). La tabla VII muestra una sección de la tabla de distancias que se ha calculado para los valores de productos y servicios presentes en cada portal.

Caso	1: ABC	2: HOY	3: VOC	4: VAN	5: MUN	6: INF	7: FAR	8: NES	9: LEV	10: 5DI
1: ABC	0,000	21,000	16,000	17,000	21,000	21,000	20,000	20,000	20,000	25,000
2: HOY	21,000	0,000	19,000	16,000	16,000	16,000	15,000	19,000	21,000	30,000
3: VOC	16,000	19,000	0,000	19,000	25,000	23,000	24,000	22,000	24,000	29,000
4: VAN	17,000	16,000	19,000	0,000	26,000	24,000	21,000	21,000	23,000	28,000
5: MUN	21,000	16,000	25,000	26,000	0,000	14,000	13,000	21,000	13,000	26,000
6: INF	21,000	16,000	23,000	24,000	14,000	0,000	7,000	13,000	19,000	22,000
7: FAR	20,000	15,000	24,000	21,000	13,000	7,000	0,000	12,000	18,000	25,000
8: NES	20,000	19,000	22,000	21,000	21,000	13,000	12,000	0,000	22,000	23,000
9: LEV	20,000	21,000	24,000	23,000	13,000	19,000	18,000	22,000	0,000	19,000
10: 5DI	25,000	30,000	29,000	28,000	26,000	22,000	25,000	23,000	19,000	0,000

Tabla VII. Sección de la matriz de distancias euclídeas establecidas entre los componentes de los portales

El siguiente paso de nuestro proceso es elegir el método de agrupamiento a seguir. Nuestra idea es reunir a los más comunes para que esos agrupamientos, de alguna manera, exhiban posibles tendencias similares de desarrollo y punteen algún tipo de disimilitudes frente a otros posibles grupos de portales. Dos son los métodos recomendables para conseguir este objetivo, el del “promedio inter-grupos” (*average linkage between groups*) y el del “vecino más cercano” (*nearest neighbour technique*). El primero de estos métodos define la distancia entre dos clusters como la media de todas las distancias posibles dos a dos de los elementos de uno y otro cluster. Este método hace uso de todos los pares de distancia, no únicamente de los

elementos más próximos o alejados, como hacen otros métodos de agrupamiento, como por ejemplo el del “vecino más cercano”, que establece la distancia para formar un *cluster* como la mínima existente entre dos elementos; una vez construido este nuevo *cluster*, se recalculan las distancias del resto de los elementos con el nuevo agrupamiento (tomando como base siempre la menor existente con alguno de los dos elementos) y se crea el siguiente agrupamiento determinando la menor medida existente.

Considerando las ventajas e inconvenientes de ambos métodos, observamos que los textos especializados en estadística se inclinan por la agrupación del “promedio inter-grupos”, más sensible a posibles transformaciones monótonas de los datos (la evolución del desarrollo de los portales sigue un ritmo que cabría considerar así); en cambio, los textos especializados en minería de datos apuestan por el uso del método del “vecino más cercano”. En realidad, ambos criterios parecen adecuados para nuestro experimento y vamos a emplear ambos con la idea de contrastar los resultados e intentar así reforzar las conclusiones que del mismo se deriven.

3. Resultados del agrupamiento por “promedio inter-grupos”

El punto final de un proceso de agrupamiento es conseguir un único *cluster* con todos los objetos de la muestra. Evidentemente el objeto de nuestra investigación es intentar discernir algunos grupos más afines entre ellos para establecer tendencias de desarrollo. Antes de proseguir con la experimentación, es preciso resolver el número de agrupamientos que deseamos obtener. Esta decisión no puede tomarse de forma arbitraria, delegando el resultado a las características propias de la muestra.

En principio, el pequeño tamaño de la muestra es un factor que debe ser considerado: no parece lógico fijar un número grande de agrupamientos porque muchos de ellos serían de un par de elementos e incluso de un elemento único. El objetivo de nuestro estudio es otro aspecto a considerar antes de fijar el número deseado de *clusters* pues de nada sirve identificar dos agrupamientos porque difícilmente se determinarían tendencias de desarrollo con esa información, especialmente si llegara a ocurrir que cada *cluster* agrupara una mitad de los elementos de la muestra, mostrando así escasa información. Tampoco sería extraño que un *cluster* recogiera un par de elementos y el otro congregase al resto de la muestra, lo que suministraría una información del tipo “destacan dos portales sobre el resto”, la cual tampoco cabría encuadrar como de “análisis de tendencias” sino más bien como de “certificación de no tendencias”. Lo más conveniente es observar cómo evolucionan las formaciones de *clusters* a partir de un número mínimo de grupos (que bien podría ser tres), hasta un número mayor de casos posibles (que en razón del tamaño de la muestra nos parece adecuado fijar en nueve, el triple del número mínimo establecido), evitando así los dos problemas planteados y extrayendo conclusiones de forma global basadas en las características de la mues-

	9	8	7	6	5	4	3
	conglomerados						
1: ABC	1	1	1	1	1	1	1
2: HOY	2	2	2	2	2	2	2
3: VOC	1	1	1	1	1	1	1
4: VAN	1	1	1	1	1	1	1
5: MUN	3	3	2	2	2	2	2
6: INF	4	2	2	2	2	2	2
7: FAR	4	2	2	2	2	2	2
8: NES	4	2	2	2	2	2	2
9: LEV	3	3	2	2	2	2	2
10: 5DI	5	4	3	3	3	3	3
11: AS	6	5	4	4	3	3	3
12: EPS	6	5	4	4	3	3	3
13: EMU	6	5	4	4	3	3	3
14: EXP	7	6	5	5	4	4	3
15: MAR	7	6	5	5	4	4	3
16: EPE	2	2	2	2	2	2	2
17: SPO	4	2	2	2	2	2	2
18: PAN	1	1	1	1	1	1	1
19: COPE	8	7	6	2	2	2	2
20: RNE	8	7	6	2	2	2	2
21: RCA	3	3	2	2	2	2	2
22: SER	8	7	6	2	2	2	2
23: CERO	8	7	6	2	2	2	2
24: 40	9	8	7	6	5	4	3
25: 100	8	7	6	2	2	2	2
26: TVE	8	7	6	2	2	2	2
27: CNN+	8	7	6	2	2	2	2
28: A3	8	7	6	2	2	2	2
29: T5	8	7	6	2	2	2	2
30: PLUS	8	7	6	2	2	2	2
31: VIL	8	7	6	2	2	2	2
32: DIR	8	7	6	2	2	2	2.

Tabla VIII. Agrupamientos/conglomerados de pertenencia de los portales periodísticos agrupados por el método “promedio inter-grupos”

tra analizada, sin fijar previamente el número de agrupamientos a obtener, acción que podría alterar el resultado o disminuir el alcance del análisis.

Si se observan detenidamente los conglomerados de pertenencia de los portales periodísticos (tabla VIII), se divisa un fuerte agrupamiento de los portales *El País*, *El Mundo* y *As*. En un principio, a estos portales les acompañan los de *5 Días* y *Expansión*, el portal de origen radio *Cadena 40* y el del diario deportivo *Marca*. A medida que aumenta el número de agrupamientos, estos últimos portales conforman agrupamientos nuevos y más pequeños entre ellos (es el caso de *Marca* y *Expansión*, que siempre van a estar en el mismo *cluster*), o bien llegan a formar *clusters* individuales (*Cadena 40* y *5 Días*), pero se van a mantener alejados del resto de los portales analizados.

Los portales *Vocento*, *ABC*, *La Vanguardia* y *Panorama Actual* se agrupan en el *cluster* 1, situación que no varía en ningún momento a medida que aumenta el número de *clusters*. De hecho, el aumento de *clusters* lo único que consigue es reducir el número de elementos que inicialmente formaban parte del *cluster* 3 anteriormente comentado. El resto de portales, que son mayoría, se agrupan en el *cluster* 2, que aglomera a los portales de diarios de prensa de grupos medianos de comunicación con un desarrollo inicial de productos y servicios y a casi todos los portales de origen radio y televisión. Este agrupamiento inicial apenas sufre alteraciones con el aumento del número de *clusters* fijados. Dentro de este agrupamiento también se encuadran los portales diseñados específicamente para Internet.

3	4	5	6	7	8	9
(agrupamiento) nº elementos	(agrupamiento) nº elementos	(agrupamiento) nº elementos	(agrupamiento) nº elementos	(agrupamiento) nº elementos	(agrupamiento) nº elementos	(agrupamiento) nº elementos
(1) 4	(1) 4	(1) 4	(1) 4	(1) 4	(1) 4	(1) 4
(2) 21	(2) 21	(2) 21	(2) 21	(2) 9 (6) 12	(2) 6 (3) 3 (7) 12	(2) 2 (3) 3 (4) 4 (8) 12
(3) 7	(3) 4 (4) 3	(3) 4 (4) 2 (5) 1	(3) 1 (4) 3 (5) 2 (6) 1	(3) 1 (4) 3 (5) 2 (7) 1	(4) 1 (5) 3 (6) 2 (8) 1	(5) 1 (6) 3 (7) 2 (8) 1

Tabla IX. Evolución del número y tamaño de los clusters de los portales

La evolución de los *clusters* demuestra que las diferencias pueden observarse con un pequeño número de *clusters* y que no es necesario llegar a establecer hasta nueve. En realidad, ningún portal cambia de agrupamiento; los grandes agrupamientos iniciales 2 y 3 se van a ir dividiendo en otros menores a medida que se establece un número mayor de particiones. Así, el agrupamiento 3 no cambia ya a partir de la cuarta columna de la tabla IX (cuando se establecen seis *clusters*) y es justo a partir de ahí cuando comienza la descomposición del *cluster* inicial 2 en

otros menores. Antes hemos indicado que un agrupamiento de un solo elemento no sirve prácticamente para nada en nuestro estudio, ya que no aporta nada en el establecimiento de tendencias comunes de desarrollo. Siguiendo esta lógica y analizando la anterior tabla de izquierda a derecha (en orden creciente de *clusters* establecidos), la última columna donde no aparece un *cluster* de elemento único es la segunda, correspondiente al establecimiento de un número de 4 agrupamientos. Esta división es la última que no nos proporciona *clusters* individuales y bien podría considerarse como el agrupamiento más adecuado para nuestro objetivo según las características de la muestra analizada y la evolución de los agrupamientos sucesivos.

El análisis de estos cuatro *clusters* proporciona datos adicionales a los aportados por los estudios anteriores. La primera conclusión que se extrae es doble: los portales de los dos principales periódicos impresos comparten agrupamiento. Esto implica que su grado de desarrollo es paralelo y vienen marcando una tendencia. Puede llamar la atención la presencia de un portal de un diario deportivo en este *cluster* con un desarrollo menor en cuanto a productos y servicios (el portal del diario *As*), aunque su presencia (que se mantiene en toda la tabla) puede venir justificada por el hecho de pertenecer al mismo grupo de comunicación que el diario *El País* (grupo PRISA). Este hecho permite extraer una nueva conclusión (ya vislumbrada en trabajos previos pero sin elementos justificantes): cuando un grupo de comunicación desarrolla un producto o servicio para uno de sus portales y verifica su aceptación por los usuarios, lo traslada al resto de sus portales.

Los portales de los diarios *Expansión* y *Marca* comparten agrupamiento con el portal de la *Cadena 40*. Los dos primeros portales pertenecen al mismo grupo de comunicación, RECOLETOS, y parecen ocupar un segundo nivel de desarrollo frente al *cluster* anterior. Con ellos, aparece el único portal de origen radio que en estudios anteriores había demostrado poseer un conjunto de características que le permitiría distinguirse del escaso nivel de desarrollo de los portales de similar tipología (Guerrero, 2003; López Carreño, 2004). En realidad, este nuevo agrupamiento no hace más que confirmar el superior nivel de desarrollo de los portales de los grupos PRISA y RECOLETOS frente al de los demás grupos. Además, permite entrever que estos dos grupos siguen una tendencia común algo alejada del resto de portales. La única diferencia entre el agrupamiento 3 y el agrupamiento 2 estriba en los niveles de desarrollo del segundo *cluster*, que parecen algo menores. Esta conclusión se confirma posteriormente al aumentar el número de *clusters* establecidos, el agrupamiento 3 sigue constante y el 2 se divide en dos (la *Cadena 40* forma un *cluster* de elemento único).

El *cluster* 1 aglomera al portal *Vocento* junto con *ABC* (ambos del mismo grupo de comunicación —VOCENTO—, circunstancia que comienza a aparecer como lógica causa de agrupamiento de portales desarrollados bajo una misma tendencia,

la de su grupo de comunicación), y a otros dos portales de grupos distintos, *La Vanguardia* y *Panorama Actual* (6). En este agrupamiento aparecen el resto de portales cuyo número de elementos supera la mitad de los posibles y parece plausible comentar que comparten las suficientes características en común como para identificar una tendencia de desarrollo con una serie de rasgos propios que lo alejan del camino marcado por PRISA y RECOLETOS. Este alejamiento se mantiene constante a medida que aumenta el número de *clusters* establecidos, lo que permite reafirmarnos en este postulado.

El *cluster 2*, el mayoritario, confirma el heterogéneo nivel de desarrollo de la mayoría de los portales periodísticos y las escasas diferencias existentes en el desarrollo de los portales de radio y televisión (que con la excepción de la *Cadena 40* aparecen todos aquí englobados). Estos portales de radio y televisión comparten además el hallarse en un nivel inicial de desarrollo y realmente no parecen introducir diferencias sustanciales con los de origen prensa de desarrollo paralelo, excepto cuando aumenta el número de *clusters* establecidos, donde se mantiene el fuerte vínculo entre los portales de radio y televisión y se detectan algunos matices que los diferencian de los de prensa. Los portales desarrollados específicamente para Internet se engloban también en este *cluster* y no presentan ningún aspecto especialmente significativo.

Un dendograma representa la evolución de los agrupamientos y determina los grupos más homogéneos presentes dentro de la muestra analizada de portales periodísticos. En un dendograma, además de estar representadas las etapas del proceso de fusión, se observan las distancias existentes entre los elementos unidos, aunque en una escala estandarizada de 25 puntos: las líneas verticales identifican elementos unidos en un mismo *cluster* y la posición de las mismas representa la distancia a la que se ha llevado a cabo el agrupamiento. Cuanto más a la derecha estén estas líneas, significa que los grupos eran más heterogéneos.

El dendograma de la figura 1 solo muestra un conglomerado muy homogéneo (el formado por los portales *RNE*, *TVE* y *Cadena 100*), curiosamente el primero que se lleva a cabo. Estos portales poseen un muy escaso nivel de desarrollo y solo disponen de los elementos básicos, lo que explica su alto grado de homogeneidad. Pocos son los conglomerados formados antes de llegar a la posición 10 de la escala, aunque a partir de ahí se van produciendo algunos de los que marcan tendencias. Por ejemplo, el conglomerado inicial (el citado anteriormente) aumenta poco a poco incorporando a casi todas los portales de radio y televisión (con excepción de *Cadena 40*) y ya se encuentra formado en la mitad del dendograma. A continuación se le unen los portales de origen Internet (aproximadamente en la posición 15 de la escala), y los de prensa regional o de grupos de comunicación (estos son más heterogéneos en conjunto, hecho lógico por su número y variedad, pero dentro de este subgrupo se han producido aglomerados antes de llegar al valor 15 de la escala).

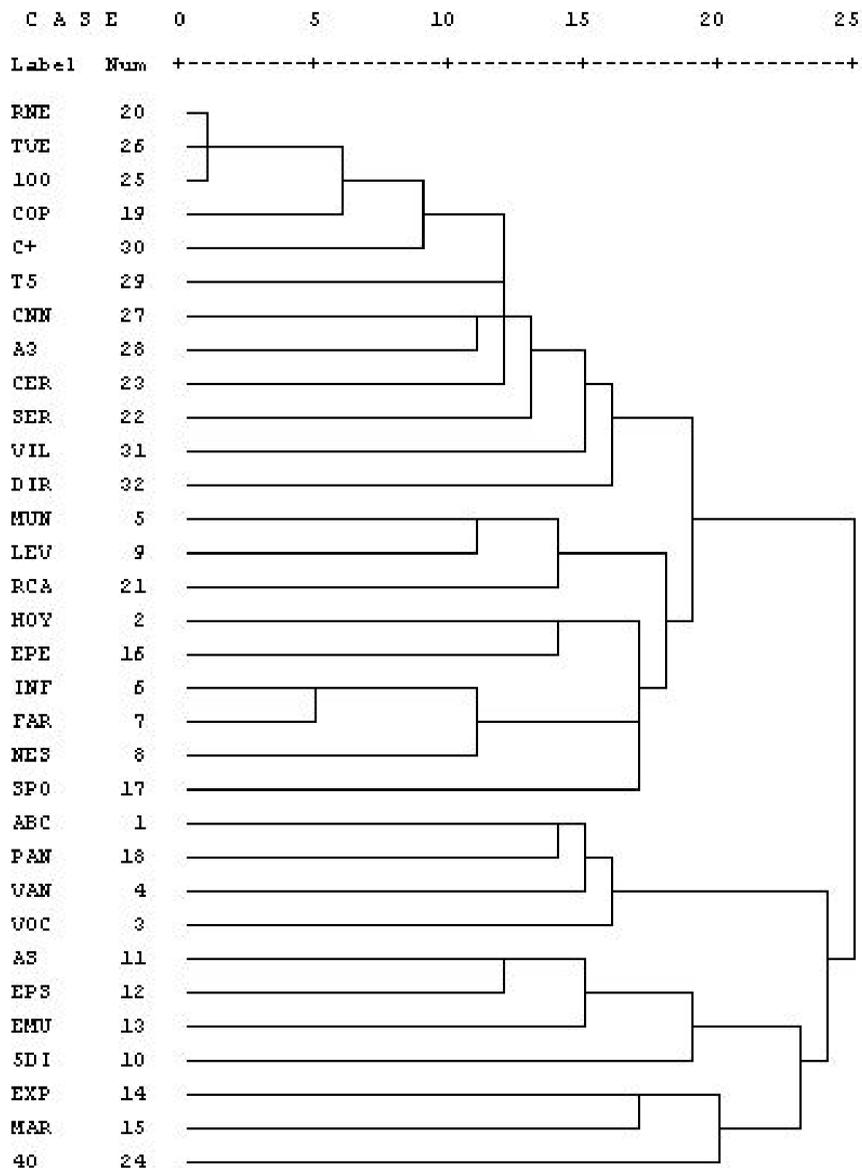


Fig. 1. Dendrograma del cluster jerárquico de los portales con el método “promedio inter-grupos”.

Este grupo ya no se une al resto de los portales (correspondientes a los *clusters* 1 y 3 de la primera columna de la tabla VIII) hasta el final del proceso. Estos portales se han aglomerado dentro de cada *cluster* en distancias algo elevadas —entre las escalas 15 y 20 para ser más exactos—, lo que revela la gran heterogeneidad en la estructura de los portales periodísticos analizados. El pertenecer a dos grupos periodísticos distintos puede justificar estos elevados niveles de distancia, ya que cada empresa tiende a adoptar los elementos más comunes en el diseño de sus portales e intenta ofrecer productos y servicios que no se encuentren en la competencia. Si nos situamos en la parte inferior del dendograma, donde figuran los portales *As*, *El País*, *El Mundo*, *5 Días*, *Expansión*, *Marca* y *Cadena 40*, observamos que las uniones entre portales del mismo grupo (*As* y *El País*, por ejemplo) se llevan a cabo antes de la mitad del dendograma, y su unión definitiva con *El Mundo* ocurre antes de la posición 15, sucediendo lo mismo un poco más adelante entre *Expansión* y *Marca*.

No hay que olvidar tampoco que, ateniéndonos estrictamente al número de sus elementos componentes, todos estos portales tienen un nivel intermedio o avanzado de desarrollo. En definitiva, el dendograma muestra que el vínculo actual que permite establecer tendencias en la implantación de estos sistemas de información reside básicamente en el alto grado de desarrollo que implica la coincidencia en los elementos más comunes. Este porcentaje de coincidencia variará con el tiempo, seguramente irá en aumento, aunque podría darse el caso (poco probable en un principio), de que los distintos grupos aplicaran líneas de desarrollo de productos y servicios muy dispares, algo que creemos no ocurrirá porque la web es un entorno muy abierto y si un nuevo servicio gana en popularidad entre los usuarios de ese portal nada tiene de extraño que los demás lo implementen rápidamente. Por ello estamos casi seguros de que con el paso del tiempo estas uniones deberían producirse un poco más a la izquierda, lo que implicaría la existencia de una mayor homogeneidad entre portales.

4. Resultados del agrupamiento por el “vecino más cercano”

En este caso analizamos el comportamiento del agrupamiento de los portales periodísticos por medio de la técnica de agrupamiento conocida como “vecino más cercano”, en la cual los emparejamientos se llevan a cabo siempre entre los dos elementos (o elementos y *clusters*), más próximos.

Esta modalidad de agrupamiento ofrece un resultado ligeramente diferente del anterior, aunque sustancialmente no aporta conclusiones contradictorias. En una primera instancia se observa que el *cluster* (*El País*, *El Mundo*, *As*) vuelve a formarse y se mantiene en todos los niveles de agrupamiento. Como novedad, el portal *5 Días* forma un *cluster* de único elemento en todos los casos e inicialmente se forma un gran *cluster* con el resto de portales (un total de 28), del cual se desgajan paso

	9	8	7	6	5	4	3
	conglomerados						
1: ABC	1	1	1	1	1	1	1
2: HOY	2	2	2	2	2	2	1
3: VOC	3	3	3	1	1	1	1
4: VAN	4	4	2	2	2	2	1
5: MUN	2	2	2	2	2	2	1
6: INF	2	2	2	2	2	2	1
7: FAR	2	2	2	2	2	2	1
8: NES	2	2	2	2	2	2	1
9: LEV	2	2	2	2	2	2	1
10: 5DI	5	5	4	3	3	3	2
11: AS	6	6	5	4	4	4	3
12: EPS	6	6	5	4	4	4	3
13: EMU	6	6	5	4	4	4	3
14: EXP	7	7	6	5	5	2	1
15: MAR	8	8	7	6	1	1	1
16: EPE	9	2	2	2	2	2	1
17: SPO	2	2	2	2	2	2	1
18: PAN	1	1	1	1	1	1	1
19: COPE	2	2	2	2	2	2	1
20: RNE	2	2	2	2	2	2	1
21: RCA	2	2	2	2	2	2	1
22: SER	2	2	2	2	2	2	1
23: CERO	2	2	2	2	2	2	1
24: 40	2	2	2	2	2	2	1
25: 100	2	2	2	2	2	2	1
26: TVE	2	2	2	2	2	2	1
27: CNN+	2	2	2	2	2	2	1
28: A3	2	2	2	2	2	2	1
29: T5	2	2	2	2	2	2	1
30: PLUS	2	2	2	2	2	2	1
31: VIL	2	2	2	2	2	2	1
32: DIR	2	2	2	2	2	2	1

Tabla X. Agrupamientos/conglomerados de pertenencia de los portales periodísticos agrupados por el método del “vecino más cercano”

a paso aquellos con un mayor nivel particular de desarrollo: *ABC* y *Panorama Actual*, *Vocento*, *La Vanguardia*, *Expansión*, *Marca* y *El Periódico* (esto ocurre cuan-
Scire. 11 : 2 (sep.-dic. 2005) 33-62. ISSN 1135-3761.

do se llegan a establecer nueve *clusters*). Los portales de origen radio, televisión e Internet permanecen juntos en todos los niveles de agrupamiento.

Quizá este método ofrece unos resultados algo más dispersos que el anterior y tiende a generar *clusters* de único elemento, pero sus resultados en líneas generales confirman los anteriores.

5. Resultados del análisis de los grupos de comunicación

Realizaremos el análisis de *clusters* jerárquicos sobre los datos de la tabla V (valores medios de presencia de elementos en portales agrupados en función de su grupo de comunicación), usando los dos métodos anteriores de agrupamiento. En la creación de la matriz de grupos de comunicación se sigue el criterio de calcular los valores medios de cada componente en un grupo. Algunos de los portales de la muestra, aquellos no adscritos a los grupos de comunicación analizados, no forman parte de esta nueva matriz (es el caso de los portales *Panorama Actual*, *Diario Directo* y *Vilaweb*). Esta matriz de grupos y elementos ha permitido la obtención de la siguiente tabla de distancias:

Caso	Distancia euclídea al cuadrado								
	1: VOCENTO	2: VANGUAR	3: IBÉRICA	4: PRISA	5: RECOLE	6: ZETA	7: COPE	8: RTVE	9: ANTENA3
1: VOCENTO	0,000	6,063	6,625	7,996	8,479	9,063	15,563	16,563	10,813
2: VANGUAR	6,063	0,000	6,563	9,827	10,167	8,750	17,000	18,750	11,500
3: IBÉRICA	6,625	6,563	0,000	8,532	12,063	8,813	12,313	14,063	10,313
4: PRISA	7,996	9,827	8,532	0,000	7,374	12,077	10,112	13,219	9,398
5: RECOLE	8,479	10,167	12,063	7,374	0,000	13,417	20,167	22,750	17,500
6: ZETA	9,063	8,750	8,813	12,077	13,417	0,000	13,250	15,500	10,750
7: COPE	15,563	17,000	12,313	10,112	20,167	13,250	0,000	4,250	8,000
8: RTVE	16,563	18,750	14,063	13,219	22,750	15,500	4,250	0,000	7,750
9: ANTENA3	10,813	11,500	10,313	9,398	17,500	10,750	8,000	7,750	0,000

Esta es una matriz de disimilaridades

Tabla XI. Representación de las distancias euclídeas de la matriz de grupos y elementos

La única variante introducida en este nuevo procedimiento de cálculo es la reducción del número máximo de *clusters* a establecer debido a que el tamaño de la muestra es mucho menor que en el caso anterior (se ha pasado de 32 elementos a 9), por ello se va a calcular un rango de 3 a 6 conglomerados. Aplicando la técnica de agrupación “promedio inter-centros” se obtienen los agrupamientos de grupos de comunicación de la tabla XII.

	6 conglomerados	5 conglomerados	4 conglomerados	3 conglomerados
1: VOCENTO	1	1	1	1
2: VANGUAR	1	1	1	1
3: IBÉRICA	1	1	1	1
4: PRISA	2	2	2	2
5: RECOLE	3	2	2	2
6: ZETA	4	3	3	1
7: COPE	5	4	4	3
8: RTVE	5	4	4	3
9: ANTENA3	6	5	4	3

Tabla XII. Evolución de los conglomerados de los grupos de comunicación con la técnica “promedio inter-grupos”

Estos agrupamientos confirman los datos anteriores sobre la tendencia marcada por los grupos PRISA y RECOLETOS, que siguen una evolución común (que solo se rompe en el caso extremo de seis conglomerados), al igual que VOCENTO y LA VANGUARDIA, a los que se une en este caso el grupo PRENSA IBÉRICA. El grupo ZETA forma un agrupamiento de un único elemento de forma permanente. Este nuevo punto de vista permite ratificar la semejanza entre los portales de origen solo radio o solo televisión (COPE, RTVE y ANTENA3), aunque introduce una novedad: en sus agrupamientos excluyen a grupos con portales de origen prensa.

Dendrogram using Average Linkage (Between Groups)

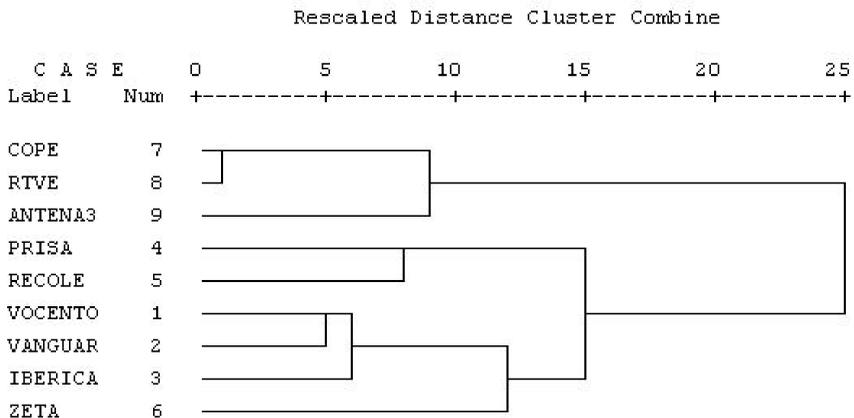


Fig. 2. Dendrograma del cluster jerárquico de los grupos de comunicación agrupados con la técnica “inter-grupos”.

La interpretación del dendograma de la figura 2 brinda una serie de datos que confirman la serie de indicios apuntados anteriormente. Los portales de grupos de comunicación dedicados exclusivamente a la radio o a la televisión (COPE, RTVE y ANTENA 3) se unen en niveles bajos, lo que ratifica su homogeneidad y escaso nivel de desarrollo (se unen tan pronto porque apenas disponen de elementos que los distinguen, solo incorporan elementos básicos en sus portales).

Un segundo resultado de los agrupamientos mostrados por el dendograma anterior es la ratificación de la gran distancia que deben recorrer los portales de estos grupos para llegar a asemejarse a los de algún otro (como se observa, la unión con el resto de portales solo se lleva a cabo en el último momento, posición 25 de la escala del dendograma). Cuando se unifican los datos de los distintos portales de cada grupo de comunicación, los vínculos se realizan en niveles más bajos, es decir, los valores medios por grupos de comunicación tienden a ser más homogéneos (circunstancia lógica si las medias son representativas, tal como parecen serlo). Así, el agrupamiento de los grupos PRISA y RECOLETOS se lleva a cabo antes de la posición 10 de la escala del dendograma, llamando la atención que LA VANGUARDIA y VOCENTO se hayan fusionado en la posición 5, y que se unan a PRENSA IBÉRICA en el paso siguiente. De hecho, la unión de todos los grupos de desarrollo avanzado o intermedio se lleva a cabo en la posición 15.

En último lugar de nuestro estudio, se va a obtener el *cluster* de los grupos de comunicación siguiendo el método del “vecino más cercano”, con el fin de observar si se introduce alguna variante significativa.

	6	5	4	3
	conglomerados	conglomerados	conglomerados	conglomerados
1: VOCENTO	1	1	1	1
2: VANGUAR	1	1	1	1
3: IBÉRICA	1	1	1	1
4: PRISA	2	2	2	1
5: RECOLE	3	2	2	1
6: ZETA	4	3	3	2
7: COPE	5	4	4	3
8: RTVE	5	4	4	3
9: ANTENA3	6	5	4	3

Tabla XIII. Evolución de los conglomerados de los grupos de comunicación con la técnica del “vecino más cercano”

Tal como era de esperar, aunque se observa una pequeña diferencia en cuanto a los agrupamientos iniciales (con tres *clusters*), el comportamiento de la evolución de los *clusters* con esta técnica en los siguientes pasos es similar al caso anterior, lo que reafirma las conclusiones anteriores.

6. Conclusiones

El análisis del conjunto de resultados que se han venido presentando en el apartado anterior permite la extracción del siguiente conjunto de conclusiones.

En lo relacionado con las tendencias de desarrollo se han conseguido identificar cuatro estilos suficientemente diferenciados: (I) el seguido por los portales correspondientes a las grandes cabeceras de prensa nacional (*El País* y *El Mundo*) con el portal del diario deportivo *As*, que combinan desarrollo avanzado con la presencia de muchos elementos comunes entre ambos; (II) un segundo grupo formado por el resto de portales de estos dos grandes grupos de comunicación (*Cadena 40*, *5 Días*, *Expansión* y *Marca*), caracterizados por un número menor de componentes pero con una línea similar a la anterior; (III) portales de menor desarrollo pertenecientes a grupos de comunicación menores que los anteriores pero de cierta importancia (*La Vanguardia*, *Vocento*, *ABC*, *Panorama Actual*), y (IV), compuesto por un amplio grupo de portales, se encuentra aún en un nivel de desarrollo muy inicial, lo que hace prácticamente imposible establecer tendencias entre ellos. Las distancias entre los grandes agrupamientos indicados anteriormente son muy altas, lo que equivale a unas grandes diferencias en la estructura de los portales analizados. Estos valores representan claramente a distintas líneas de trabajo de los gestores de los portales. Puede ocurrir que un portal opte por desarrollar servicios de valor añadido (es el caso del portal *El País*) y otro encamine su oferta hacia los productos documentales (muy presentes en el portal *Vocento*, por ejemplo). Esto provoca mayor distancia que la observada a primera vista, cuando simplemente se totaliza la oferta de elementos componentes del portal. Dentro de estos grandes agrupamientos las fusiones también se realizan a una gran distancia según la escala de distancias indicada por el dendograma de agrupamientos, excepto cuando el nivel de desarrollo de los portales es ínfimo (como ocurre con los portales de origen radio y televisión), y las fusiones se llevan a cabo en los primeros pasos del algoritmo de agrupamiento. Estas distancias deben descender a medida que los portales vayan incorporando servicios de valor añadido entre sus componentes y asemejen un poco más sus estructuras. De hecho, aquellos que los incorporan en la actualidad se encuentran más cercanos.

Los portales de origen prensa están mucho más desarrollados y forman parte de las tendencias en evolución en mayor medida que los portales de origen radio, televisión o Internet. Se confirma, por otra parte, que no existen diferencias significativas entre los portales de origen radio y televisión. Los de origen Internet no han aportado ninguna consideración especial en este experimento. De hecho, han pasado prácticamente desapercibidos. Al estudiar los grandes grupos de comunicación se reducen las distancias entre los objetos de la muestra. Los grupos PRISA y RECOLETOS parecen marcar la tendencia de desarrollo de portales por grupo de comunicación más completa, seguida por el resto de grupos de comunicación

(VOCENTO, LA VANGUARDIA, PRENSA IBÉRICA y ZETA), quedando algo relegados en este caso los portales de grupos de comunicación específicos de radio o televisión.

A título de reflexiones finales se puede apuntar que estos resultados muestran una realidad bastante heterogénea por dos razones: la primera es que el espacio analizado es diverso (con lo cual los resultados no hacen más que reflejar la realidad), y la segunda tiene que ver más con la cierta novedad de la presencia de muchos de estos medios de comunicación en la web. Si bien algunos de estos portales llevan muchos años en la web, el tiempo que la mayoría llevan funcionando bajo el formato de portal es menor y los niveles de desarrollo se ven tan influenciados por la potencia del grupo que respalda al portal como por la calidad de su diseño. Estamos seguros de que revisando estos datos en un breve plazo se obtendría una mayor uniformidad en los diseños de estos portales, ya que los que aún están en fase de inicio incorporarán lógicamente elementos ya desarrollados por la competencia, aunque, claro está, los que van por delante también seguirán introduciendo nuevos elementos en esas estructuras.

Lo que sí es incuestionable es que la realización de un estudio de esta naturaleza representa un salto cualitativo en el análisis de las estructuras de los portales periodísticos, ya que permite corroborar otros trabajos previamente realizados y abre nuevas vías de estudio, tanto dentro de este tema como dentro de la propia web en general.

7. Perspectivas futuras

Son variadas las nuevas vías complementarias de investigación que pueden desarrollarse a partir de lo realizado en este trabajo. La primera y más inmediata ha sido apuntada varias veces. Este mismo método de estudio e investigación puede aplicarse con ligeras modificaciones en el proceso de sistematización de los elementos componentes a otra serie de portales de la web (generalistas, de ISP's, de banca electrónica, de comercio electrónico, etc.). Estos procesos de revisiones pierden valor si no se repiten en breves periodos de tiempo. Por ello resultaría apropiado disponer de un medio dinámico de comunicación con los gestores de contenidos de los portales periodísticos, con el fin de que ellos participaran en el refresco y actualización de los datos correspondientes a sus portales. Quizá la puesta en marcha de un sitio web a modo de "observatorio" podría contribuir a este tema.

Lo más lógico es que los portales periodísticos completen en breve plazo su oferta de productos informativos. Esto equivale a decir que las nuevas investigaciones de tendencias estarán inevitablemente orientadas a la oferta de productos documentales (aún muy escasa) y de servicios de valor añadido (donde solo un reducido número de portales viene actualmente trabajando).

Otro enfoque para estas investigaciones podrá ser analizar qué productos o servicios discriminan más a un portal de otro, o a un grupo de comunicación de otro, o bien cómo se posicionan los portales en torno a un determinado elemento. Para este tipo de análisis se pueden emplear las técnicas de descomposición singular de valores —que también son empleadas en los procesos de indización por semántica latente (Berry, Dumais y O'Brien, 1995; Letsche y Berry, 1996; Rosario, 2000) y han sido recientemente introducidas en procesos de minería de datos (7)— para proceder al filtrado de esta información.

Al igual que las técnicas de *clustering*, la indización por semántica latente también trabaja con las distancias, por lo que es lógico suponer que puede ser empleada también para determinar tendencias y patrones. Paradójicamente, para que los algoritmos funcionen adecuadamente y ofrezcan resultados verdaderamente destacables, la matriz a simplificar inicialmente ha de ser de considerables proporciones. En este caso, no se trata solo del número de objetos de la muestra, sino también del conjunto de fenómenos observables en cada portal (sus elementos). El número de 60 elementos actualmente desarrollado no puede considerarse alto, por lo que cuantos más elementos se distingan en la estructura de los portales más cerca estaremos de verificar si este método puede ser una alternativa para la realización de estos estudios. Así resultará necesario también promover nuevas revisiones exhaustivas de los elementos componentes de los portales periodísticos.

Notas

- (1) Los datos de la consultora Nielsen Net Ratings (2004) indican que los sitios más visitados en España el último mes de febrero son *Microsoft*, *Google*, *Terra*, *Wanadoo*, *Yahoo!*, *AOL Time Warner* y *T-Online*, todos portales a excepción de *Google* (en pleno proceso de convertirse en portal). La misma fuente indica que en febrero de 2004, en el Reino Unido los sitios más visitados son *Microsoft*, *Google*, *Yahoo!*, *eBay*, *BBC* y *AOL Time Warner*. En los Estados Unidos y en la misma fecha, los sitios más visitados desde el hogar eran *Microsoft*, *AOL Time Warner*, *Yahoo!*, *Google* y *eBay*.
- (2) El Estudio General de Medios puede consultarse en el sitio web de la Asociación para la Investigación de los Medios de Comunicación. <<http://www.aimc.es/>> . Consulta: 8 septiembre 2005.
- (3) Realmente la muestra representa un número mayor de portales ya que algunos grupos de comunicación han desarrollado modelos genéricos de portal que aplican a la mayoría de sus portales periodísticos, tal como ocurre con el grupo VOCENTO (se ha analizado el portal del diario *La Verdad* como representante de los portales de todos los diarios regionales de este grupo) y el grupo ZETA (se ha analizado *El Periódico de Catalunya*). También se da el caso de que un mismo grupo emplea varios modelos de portal; entonces se analizan todos por separado (curiosamente también sirve como ejemplo el grupo VOCENTO, que junto al genérico emplea un diseño específico para el portal del diario *ABC* y otro para el portal de *Tele 5*).

- (4) Para una explicación más detallada de los productos informativos y documentales de la tabla II se recomienda la consulta del capítulo 3 de López Carreño (2003)
- (5) Al igual que en la referencia anterior, para ampliar información sobre los servicios de valor añadido presentes en los portales periodísticos, se recomienda la lectura del mismo capítulo de la citada tesis doctoral.
- (6) El portal del diario *Panorama Actual* es el único de los de alcance regional pertenecientes a pequeños grupos de comunicación que posee un nivel medio de productos y servicios desarrollados (Martínez Rodríguez, 2003).
- (7) Sergey Brin, codirector de *Google*, hacía público en febrero de 2004 que este motor de búsqueda ha pasado a incorporar procedimientos basados en la indexación por semántica latente para solucionar gran parte de los problemas derivados del ingente volumen de su base de datos y del *web spamming*. Aunque el procedimiento es algo complejo de explicar, la idea general es reducir un espacio n-dimensional a un espacio bivariante de vectores; ello se hace posible en matrices donde la mayor parte de los valores contenidos en los vectores son nulos (tal como es el caso de las matrices de términos y documentos de un SRI). Entonces resulta sencillo reducir la matriz original a una matriz k-dimensionada en función de nuestro objeto de estudio que nos permita extraer una información muy difícil de discernir en la voluminosa matriz original.

Referencias

- Baeza-Yates, R.; Ribeiro-Neto, B. (1999). *Modern information retrieval*. Nueva York: ACM Press; Harlow, etc.]: Addison-Wesley, 1999.
- Berry, M.W., Dumais, S. T.; O' Brien, G. W. (1995). Using linear algebra for Intelligent Information Retrieval. // *SIAM Review* 37:4 (1995) 573-595.
- Berry, M.W. (1996). *Large-Scale Information Retrieval with Latent Semantic Indexing*. Knoxville, University of Tennessee, 1996. URL: <<http://www.cs.utk.edu/~berry/lsi++/>>. Consultado: 2004-03-12.
- Deuze, M. (2001). Online Journalism: Modelling the First Generation of News Media on the World Wide Web. // *First Monday*, volume 6, number 10, October 2001. URL: <http://firstmonday.org/issues/issue6_10/deuze/>. Consultado: 2004-03-12.
- Díaz Noci, J. (2001). *La escritura digital: hipertexto y construcción del discurso informativo en el periodismo electrónico*. Bilbao: Universidad del País Vasco, Servicio Editorial, 2001.
- Donaghy, R. (2002). Evaluating online newspapers using established web design. Chapel Hill, University North Carolina, 2002. URL: <<http://neoref.ils.unc.edu/2776.pdf>>. Consultado: 2004-04-01.
- Flores Vivar, J.; Arruti, A. M. (2001). *Ciberperiodismo: nuevos enfoques, conceptos y profesiones emergentes en el mundo infodigital*. México D. F.: LIMUSA, 2001.
- Fuentes i Pujol, E.; Jiménez, M. A.; González, A. (2002). El profesional de la información ante el reto de los medios de comunicación digitales: el caso de la prensa. // *Scire: Representación y Organización del Conocimiento* 8:2 (julio-diciembre 2002) 141-152.
- García Gómez, J. C. (2004). Usabilidad de las páginas de inicio de los diarios digitales españoles. // *Scire: Representación y Organización del Conocimiento* 10:2 (2004) 123-136.

- Gondar Nores, E. (2002). Introducción al Análisis Multivariante. Madrid. Data Mining Institute, 1997. URL: <<http://www.estadistico.com/arts.html?20020527>>. Consultado: 2004-03-20.
- Guerrero García, F. (2003). Análisis de los portales periodísticos de origen radio. Murcia: Facultad Comunicación y Documentación, 2003. Inédito, proyecto fin de carrera. Directores: Francisco Javier Martínez Méndez y Rosana López Carreño.
- Harper, C. (1998). Journalism in a Digital Age. MIT Communications Forum, 1998. URL: <<http://web.mit.edu/comm-forum/papers/harper.html>>. Consultado: 2004-04-08.
- Jardine, N.; Rijsbergen, C. J. van (1971). The use of hierarchical clustering in information retrieval. // Information Storage and Retrieval 7 (1971) 217-440.
- Jiménez, M. A.; Fuentes i Pujol, E.; González, A. (2000) Documentación e información electrónica. // Moreiro González, J. A. (coord.). Manual de Documentación Periodística. Madrid: Cátedra, 2000, 344-425.
- Kenney, K.; Gorelik, A.; Mwangi, S. (2000). Interactive features of Online Newspapers. // *First Monday*, 5:1 (January), 2000. URL: <http://firstmonday.org/issues/issue5_1/kenney/>. Consultado: 2004-02-20.
- López Carreño, R.; Martínez Méndez, F. J. (2002). La importancia de los productos documentales dentro de los 'e-contents' de los servicios de prensa digital. // Actas de CALSI 2002. Valencia: Universidad Politécnica y CALSI, 2002. CD-ROM.
- López Carreño, R.; Martínez Méndez, F. J. (2003). Realidad heterogénea de los productos documentales en el periodismo electrónico español. // Boletín Asociación Hispana de Documentalistas en Internet. Vol. 3. Madrid: AHDI, 2003. URL: <<http://www.documentalistas.com/AHDI/2003>>. Consultado: 2003-12-28.
- López Carreño, R. (2003). Análisis de los portales periodísticos españoles. Taxonomía de sus elementos componentes. Tesis doctoral. Alicante: Biblioteca Miguel de Cervantes, 2003. URL: <<http://www.cervantesvirtual.com/FichaObra.html?Ref=10572>>. Consultado: 2004-04-02.
- López Carreño, R. (2004). Grado de desarrollo de los portales periodísticos españoles de radio y televisión. // Scire 10:2 (2004) 115-122.
- McAdams, M. (1995). Inventing an online newspaper. // *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*, volume 3, number 3 (July). 64-90. URL: <<http://www.sentex.net/~mmcadams/invent.html>>. Consultado: 2004-02-20.
- Martínez Rodríguez, Lara. (2003) Análisis de los portales periodísticos de ámbito regional y local. Murcia: Facultad Comunicación y Documentación, 2003. Proyecto fin de carrera inédito. Directores: Francisco Javier Martínez Méndez y Rosana López Carreño.
- Medialabs (2002). Estudio sobre la presencia de los diarios españoles en la red: diciembre de 2002. Madrid: Medialabs, 2002. URL: <<http://www.medialabs.es/servicios/diarios.htm>>. Consultado: 2004-02-02.
- Moreiro González, J. A. (coord.). (2000). Manual de documentación informativa. Madrid: Cátedra, 2000.
- Nielsen, J. (1993). Usability Engineering. Academic Press: San Diego, CA, 1993.
- Nielsen Net Ratings* (<<http://www.nielsen-netratings.com/>>). Consultado: 2004-03-08.
- Scire*. 11 : 2 (sep.-dic. 2005) 33-62. ISSN 1135-3761.

- Palazón Meseguer, A. (2001) Comunicación web: el valor de los contenidos de la Red. // Comunicar: Revista Científica de Comunicación y Educación 17 (2001) 93-96.
- Peñafiel Saiz, C.; López Vidales, N. (2002) Claves para la era digital: evolución hacia nuevos medios, nuevos lenguajes y nuevos servicios. Bilbao: Universidad del País Vasco, Servicio Editorial, 2002.
- Poblaciones Magro, María (2003). Establecimiento de los componentes básicos de los portales periodísticos de las televisiones españolas. Murcia: Facultad Comunicación y Documentación, 2003. Proyecto fin de carrera inédito. Directores: Francisco Javier Martínez Méndez y Rosana López Carreño.
- Rijsbergen, C. J. (1999) Information Retrieval. Glasgow, University, 1999.
URL: <<http://www.dcs.gla.ac.uk/~iain/keith/>>. Consultado: 2001-10-21.
- Rosario, Barbara (2000). Latent Semantic Indexing: an overview. INFOSYS 240 Final Paper. Berkeley, UCLA, 2000. URL: <<http://www.sims.berkeley.edu/~rosario/projects/LSI.pdf>>. Consultado: 2004-04-07.
- Sánchez Ovcharóv, S. (2001). ¿Qué es Data Mining? Madrid: Data Mining Institute S. L., 1997.
URL: <<http://www.estadistico.com/arts.html?20010219>>. Consultado: 2004-03-18.
- Thearling, K. (2001). An Introduction to Data Mining: Discovering hidden value in your data warehouse. Thearling.com, 2004. URL: <<http://www.thearling.com/text/dmwhite/dmwhite.htm>>. Consultado: 2004-04-05.
- Tombras, Villa, R.; Rijsbergen, C. J. (2002). The effectiveness of query-specific hierarchic clustering in information retrieval. // Information Processing and Management, 38 (2002) 559-582.
- Valenzuela, J. L.; Sallent, O. (2000). WAP: las bases de Internet móvil. // Mundo Electrónico 313 (oct. 2000) 72-76. URL: <http://www.mundo-electronico.com/PDF/Any2000/313_octubre/WAP.pdf>. Consultado: 2002-10-05].
- Visauta Vinacua, B.; Martori i Cañas, J. C. (2003). Análisis estadístico con SPSS para Windows. Vol. II. Estadística Multivariante. Madrid: McGraw-Hill, 2003.