

Consecuencias de Internet para el estudio del Análisis Documental y de los lenguajes documentales: estudio de caso sobre el discurso periodístico

Antonio García Jiménez
Universidad Europea de Madrid-CEES

0.1. Resumen

Reflexión sobre la implantación de Internet como marco tecnológico para el Análisis Documental y los lenguajes documentales, en particular en el ámbito del discurso periodístico. A tal efecto se exponen los cambios producidos en el contexto teórico: el modelo de recuperación de información, el hiperdocumento como base lógico-física, la nueva configuración del discurso periodístico, las modificaciones en el trabajo del documentalista, el usuario y el Centro de Información. Finalmente se da cuenta de las nuevas bases, procedimientos y objetivos de ambos elementos con vistas a la mejora de la Organización y Representación del Conocimiento. (Autor)

Palabras clave: Análisis Documental. Lenguaje documental. Internet. Discurso Periodístico.

0.2. Abstract

Reflexion on the irruption of Internet as a technological frame of reference for "Documentary Analysis" and documentary languages, in particular in the field of journalistic discourse. To this end, changes in theoretical context are exposed: the information retrieval model, the hyperdocument as logical-physical basis, the new configuration of journalistic documents, changes in the information professional's work, the user and the Infomation Center. Finally, this paper reports on the new bases, methods and purposes of "Documentary Analysis" and documentary languages in order to improve Knowledge Organization and Representation. (Author)

Keywords: Indexing. Abstracting (Documentary Analysis). Documentary language. Internet. Journalistic Discourse.

1. Introducción

El objetivo de este estudio es plantear, desde un prisma global, algunos de los interrogantes, principios y líneas de trabajo que nacen de la implantación del Análisis Documental y de los lenguajes documentales en el marco de Internet. Factores como la posibilidad de distribución y recepción de información a nivel mundial, la inmediatez, la versatilidad y potencial personalización de formatos y de contenidos, así como la comodidad y facilidad que ofrece para el consumo *exigen el estudio de este nuevo fenómeno*. Interesa en particular el discurso periodístico no sólo por su especial configuración sino también por la relevancia que ha tenido en Internet y su consiguiente influencia social. Ahora bien, aunque nos centremos fundamentalmente en la World Wide Web, abordaremos colateralmente y de forma implícita el resto de recursos y aplicaciones.

A sabiendas de que es probable que Internet complique aún más la compleja situación paradigmática en la que se encuentra la Documentación (García Marco, 1995), nuestro principal objetivo es establecer un diagnóstico sobre la situación planteada en este nuevo dominio, las posibilidades y perspectivas que ofrece, su configuración y las nuevas vías de investigación que se plantean.

La perspectiva desde la que partimos se preocupa por los contenidos y se sitúa en contra de la coincidencia de significantes como la base del Análisis Documental, casi siempre desconectado del discurso en el que se trabaja y de los intereses de sus actantes. Del mismo modo, entendemos pertinente la superación del actual estado de los lenguajes documentales, sobre todo los combinatorios, en la medida en que deben también responder al universo en el que están inmersos.

Asimismo, creemos que para el estudio de ambos campos se debe analizar su marco teórico-práctico, formado por:

- a) La recuperación de información, como factor que determina el resto del proceso documental.
- b) El nuevo formato de los documentos.
- c) La configuración del discurso, en este caso el periodístico, puesto que el modo en que se organiza y estructura implica algunas particularidades.
- d) El propio documentalista.
- e) *El usuario*.
- f) El centro de información y el fondo documental.

Para finalizar, subrayamos que esta propuesta se encuentra inmersa en una línea de investigación que emplea, para la consecución de los objetivos docu-

mentales, el método empírico junto al recorte epistemológico sobre aquellas disciplinas que puedan aportar fundamentos y procedimientos a la Representación y Organización del Conocimiento.

2. El contexto

2.1. La recuperación de información

El primer aspecto que debe ser analizado es el de la recuperación de información, eje sobre el que gira el proceso documental. Como cuestiones relevantes destacan:

- 1) La definición de un *nuevo modelo de recuperación de información*, caracterizado por la posibilidad de navegación que la red permite a través de nodos y lazos (enlaces), formándose relaciones de tipo semántico y/o cognitivo del tal modo que se ha defendido el modelo cognitivo como el más adecuado para resolver los problemas generados (Codina, 1995).
- 2) El estudio de los *robots de búsqueda*, actualmente considerados como la gran herramienta de recuperación de información en Internet y que, en un futuro cercano, pueden ser extrapolados a entornos más restringidos. De los servidores de búsqueda —nos referimos a aquellos que buscan documentos— hay que hacer especial incidencia en el motor de análisis y en el sistema de consulta o lenguaje de interrogación. En cuanto al primer aspecto, se debe profundizar, por un lado, en los criterios y procesos de lectura, análisis, indización y resumen que se llevan a cabo, en la mayoría de las ocasiones de forma automática y, por otro lado, en la cobertura sobre las aplicaciones y los recursos, sin olvidar los distintos niveles de profundidad de los nodos y enlaces. En lo que atañe al lenguaje de interrogación, también la literatura ha señalado la falta de versatilidad dada su incapacidad para permitir que el usuario exprese de forma satisfactoria sus necesidades de información (Codina, 1997b). En realidad, el problema sigue siendo el diálogo entre el documentalista, el usuario y las máquinas.

A nuestro juicio, se debe estudiar la incorporación del factor humano, bien como complemento a la acción automatizada, bien como patrón de funcionamiento y programación a partir siempre de su anterior objetivación. En definitiva, hay que acercarse a los intereses del usuario mediante la naturalización de los procesos de búsqueda y, al mismo tiempo, prestando atención a la relevancia de los documentos, la validez de los procesos y de los enlaces, la amigabilidad de los interfaces, la rapidez en la respuesta, su claridad, la actualización y la cobertura sobre recursos,

temas y estructuras de datos (Dong y Su, 1997, p. 79). También hay que considerar la puesta en funcionamiento de la indización, el resumen y las nuevas fórmulas de interrogación junto a los procedimientos tradicionales de búsqueda de información.

- 3) El tercer aspecto que está presente es el de la *información pasiva*. En efecto, en la actualidad ya se está trabajando en programas como los agentes, las alertas y los salvadores de pantallas activos. Los primeros son programas lanzados a la red a partir de una necesidad previamente definida. Las alertas, que son perfiles definidos en bases de datos tradicionales, permiten la recepción de información en el buzón de correo electrónico (p. ej. *Business Alert* de Reuters). Finalmente están los salvadores de pantalla activos que aprovechan los momentos en que no se utiliza el ordenador para conectarse a la red y bajarse la información sobre los temas definidos previamente (Cornellá, 1996, p. 68-69).

2.2. El documento y el hiperdocumento

El segundo elemento de interés es el hiperdocumento o documento electrónico. Para realizar futuras investigaciones se tendrá que tener en cuenta su diseño y su construcción, ya sea en la identificación de los nodos que, por el momento, responde a motivaciones arbitrarias, ya sea en la creación de enlaces o lazos que los interconectan. Es necesario plantearse una metodología de diseño y una arquitectura de referencia conceptual que mejore los procesos de recuperación, la incentivación y guía del usuario, y que permita la presencia simultánea de componentes textuales, sonoros e icónicos.

Por tanto, el documentalista y el documentólogo no sólo han de conocer su base física y su base lógica (formada por el sistema de documentos y por el propio hiperdocumento) sino las formas de representación y navegación que pueden ir desde los mapas conceptuales hasta los lenguajes clasificatorios y los lenguajes facetados (Codina, 1995, p. 225). Es decir, interesan la cantidad, el tipo y las formas de acceso a la información (por ejemplo mediante encabezamientos, reenvíos, o alguna clase de índice y/o sumario) así como las formas de relación entre los documentos, bien a partir de combinaciones de palabras, bien ordenados por categorías temáticas, bien a través de redes conceptuales, etc.

En realidad, preocupa saber hasta qué punto el hipertexto refleja la forma en que el ser humano asocia la información y la busca, lo que exigiría acudir a estudios sobre la cognición y la percepción para mejorar el procesamiento de la información en el sentido que tiene en Documentación.

Así, es necesario redefinir el concepto de documento y en particular de documento electrónico. Sabemos que éste es fácilmente manipulable, enlazable inter-

namente y externamente, transformable en diversos formatos, que está dispuesto para la búsqueda de contenidos, transportable inmediatamente mediante redes de trabajo, y que su distribución es ilimitada. Características que, al fin y al cabo, dificultan tanto la descripción, el almacenamiento y la recuperación de documentos, al mismo tiempo que plantean nuevos retos a los profesionales e investigadores. De hecho, ya se han formulado nuevas nociones de documento: por ejemplo, la situada en la órbita del usuario o en el contexto a partir de una búsqueda o en términos del grado en que la información contribuye a un proyecto, a una investigación (Schamber, 1996).

2.3. Configuración del discurso: el discurso periodístico

Posiblemente sea el discurso periodístico uno de los aspectos que más incógnitas esté generando. En primer lugar, es la propia idea de discurso periodístico la que debe ser confirmada. Esta idea presupone la existencia de un conjunto de textos que tienen una reglas y estilos propios y que se articulan en un conjunto de relatos. En segundo lugar, llama la atención tanto los cambios que se han producido en los documentos periodísticos tradicionales en función de las nuevas tecnologías como el hecho de que, en el marco de Internet, los “productores de información” se inclinen más por la simple producción de información que por la elaboración de periódicos. Incluso el nuevo valor de la información no está ya situado tanto en la información generada sino en la comparación con otras fuentes de análisis y de investigación, así como en los servicios complementarios ofrecidos al usuario. En tercer lugar y sin entrar en consideraciones sobre la estructura de base de datos más conveniente, las hemerotecas digitales deben contar con el aumento de la tipología de usuarios. No sólo un periodista puede obtener información, sino también un médico podrá conocer los precedentes de un determinado asunto, o un hombre de negocios podrá analizar un proyecto comercial concreto, etc. (Aramburu y Berlanga, 1997).

En la Sociedad de la Información parece que el periodismo puede perder su preponderancia frente a otros tipos de comunicación que se están desarrollando. Surgen nuevas fuentes de información “periodística”, nuevos interrogantes acerca de la consideración de noticia o de la forma en que las personas las seleccionan en el medio electrónico. Asimismo, añadamos el aumento de las herramientas para la construcción de historias de actualidad (sonido, imágenes, texto) y las implicaciones que tiene este fenómeno tanto para la comprensión de la realidad como para la recuperación de información y la navegación. Probablemente, los cambios tendrán lugar, como ocurre en el caso del Washington Post, en torno a la actualización de los contenidos, la no coincidencia con el formato papel, la creación de nuevas secciones, el aumento de los accesos y puentes a otras fuentes de información, sin desdeñar la diversificación de las materias y de los niveles de

profundidad (Tramullas, 1996; Fulton, 1996). Como una consecuencia evidente para la investigación destaca la dificultad en la elección de unidades de trabajo.

En última instancia, se debe decidir cuál es el modelo lógico de los documentos electrónicos periodísticos, que estará casi siempre vinculado al modelo de hemeroteca o base de datos. Así pues, habrá que tener en cuenta el control sobre la estructura de los documentos conforme a los diferentes tipos de medios, las relaciones entre documentos, el control sobre la presentación, la capacidad de actualización, la forma de distribución y el formato de los denominados metadatos (Scarela et al., 1997, p. 599).

2.4. El documentalista

El rol del documentalista se verá modificado. Desde un punto de vista teórico, se debe valorar su papel como constructor de hiperdocumentos o, en su caso, como lector/productor especial que tiene por finalidad mejorar los procesos de recuperación y navegación. Es posible que Internet introduzca un nuevo paradigma, un cambio para el trabajo del documentalista, que ya se había iniciado con motivo de la aplicación documental de los sistemas expertos. En efecto, ya se han señalado las funciones del documentalista en este contexto precipitado por los procedimientos automáticos, aunque por el momento se mantendrán junto a las técnicas tradicionales (García Gutiérrez, 1996, p. 80):

- Selección de noticias y ubicación en relatos, asistencia en la normalización de términos y estructuras no reconocibles para el ordenador, actualización de productos y servicios.
- Asistencia al usuario para optimizar el diálogo con el sistema de información y realización de estudios de usuarios. Depuración de los elementos mínimos para la base de conocimientos.
- Trabajo con programas de creación de documentos electrónicos, ayudas a la navegación, colaboración con agentes automáticos.

2.5. El usuario

La globalización que supone la red, el crecimiento de las variables de trabajo, las características de los nuevos documentos y la imbricación de varios niveles de recuperación de información implican modificaciones para el estudio del usuario. En efecto, no sólo aumenta la tipología de usuarios sino la complejidad inherente a su estudio; al desbordamiento cognitivo que se puede producir en la navegación a través de la información y a la ingente cantidad de información a la que se puede llegar, se une la cantidad de decisiones que potencialmente deben ser parametrizadas en el trabajo en la red.

A este problema se suma que la figura del usuario no ha sido estudiada con la profusión ni, en la mayoría de las ocasiones, considerando todas las variables. En Internet, la mayor capacidad de decisión del usuario —hecho que se demuestra con los periódicos personalizados— aumenta los parámetros necesarios para su estudio. Finalmente, insistimos en la necesidad de conocer la información que desea el usuario y de su expresión, sin ajustarse exclusivamente a criterios economicistas sino también a criterios que den valor a la evolución cognitiva. Es obligado recordar la especificidad que caracteriza el trabajo documental dentro del discurso periodístico, y la indudable mejora que supondría desprenderse del usuario especulado y, en la medida de nuestras posibilidades, profundizar en el *usuario empírico como base para el procesamiento del lenguaje y para el buen funcionamiento de los sistemas de interrogación.*

2.6. El Centro de Información y el fondo documental

Otro de los apartados es la aparición de hemerotecas digitales, que deberán ser analizadas en función de:

- a) La definición de un modelo de organización, mediante el que los distintos componentes se relacionan y ofrecen diversos niveles de variación.
- b) La capacidad de formulación de todos los documentos periódicos.
- c) El uso de un poderoso lenguaje de interrogación que haga posible, por ejemplo, el seguimiento exhaustivo de un tema por parte de un usuario.

Así pues, deben ser puestos en funcionamiento mecanismos que permitan localizar y relacionar documentos en términos de sus propiedades y comportamientos, dentro de los periódicos (considerados como compendios de otros documentos con un conjunto de contenidos concretos) o la opción de organizar los documentos en virtud de los puntos de vista y los potenciales roles que pueden ocupar (Aramburu y Berlanga, 1997).

Al mismo tiempo, la gran cantidad de documentos existentes en la red permite contemplar las bases de datos que aquí nos preocupan como auténticos textos. De hecho, indicado su comportamiento discursivo, es viable aplicar a las bases de datos conceptos y mecanismos procedentes del análisis de contenido y del discurso, en tres dimensiones: semántica, sintáctica y de articulación intertextual (García Gutiérrez, 1996, p. 33-34).

3. El Análisis Documental

La primera cuestión que sale a relucir es la de las posibles modificaciones surgidas a la luz de la extrapolación no sólo de la concepción tradicional del Análisis Documental sino de las últimas aportaciones. En efecto, habrá que precisar cuál es la adaptación de la indización y resumen en un nuevo contexto

dominado por la incorporación de la imagen y el sonido simultáneamente, así como por el aumento de forma caótica y exponencial de la cantidad de información y donde, por ejemplo, los índices se podrían entender como herramientas de navegación. Asimismo, queda por tratar la cuestión de las variables del Análisis Documental en el discurso periodístico, a saber: la extensión, el estilo, el género, la estructura, (de las que ya se ha dado cuenta indirectamente al hablar de la nueva disposición del discurso periodístico) la ideología y los nuevos productos documentales, sobre las que será oportuno seguir investigando con vistas a una mayor fundamentación de los futuros trabajos. Tampoco se ha de desdeñar la aplicación de las nuevas propuestas de métodos de lectura documental que han supuesto un intento de objetivar los procesos del Análisis Documental (García Gutiérrez, 1992; Pinto y Gálvez, 1996).

Al mismo tiempo, entendemos que forma parte de su objeto de estudio la creación de los enlaces y la determinación de los nodos en los hiperdocumentos y su evaluación. Si uno de los objetivos del Análisis Documental es facilitar la localización de información mediante productos que representen el contenido de los documentos, también debe ocuparse de precisar, de modo objetivo y objetivado, las conexiones de contenido “semánticamente cercano” que se producen al seleccionar las ligaduras. En resumen, el documentalista tendría más que nunca la posibilidad de intervenir en la propuestas de organización de los documentos, lo que podría corresponderse, en términos de van Dijk (1990), con las superestructuras o categorías propias del discurso periodístico y que definen las formas en que los asuntos se insertan en el texto real.

4. Los lenguajes documentales

Respecto de la situación de los lenguajes documentales, y en su relación con el hipertexto, se considera que, unidos por el problema de la optimización de la representación y organización del conocimiento, ambos campos requieren una base teórica sobre cómo los individuos adquieren y representan la información (García Marco, 1994). Hasta el momento los lenguajes documentales se han empleado como base de conocimientos para la identificación automática de nodos y enlaces (lazos) o, en el apartado de las clasificaciones, han servido de modelo teórico-técnico para la construcción de algunos motores de búsqueda. En ambos casos, es patente la necesidad de una mayor reflexión documentológica.

Partiendo del papel que pueden desempeñar los lenguajes en la mejora de los procesos documentales, se ha de precisar cuál es el tipo más adecuado para el control del vocabulario y, en general, del conocimiento, si bien estamos de acuerdo con la superación de los lenguajes tradicionales mediante nuevas fórmulas de organización (en el caso de los combinatorios). De hecho, se han propuesto modelos de estructuración conceptuales que, con el objetivo de la evaluación de

los lenguajes, se han convertido en verdaderas formas estructurantes (García Jiménez, 1995). De todas formas, es previsible que, a la larga, se hagan complementarias las diferentes modalidades (jerárquicas, asociativas) y de presentación. No obstante, este no es el único factor digno de mención. Así, explotar nuevos o mejorados modos de aplicación se convierte en un objetivo deseable.

En este sentido también queda por analizar el comportamiento de las variables de los lenguajes documentales en el discurso periodístico: la temática, la base léxica, el uso y las demandas documentales, las relaciones semánticas, la muestra y la entropía. El objetivo es que la tecnología hipermedia haga más énfasis en la necesidad de conexión con los mapas cognitivos y con la búsqueda de información por parte de los usuarios. La mundialización de las redes provoca la necesidad de fórmulas de organización y control del conocimiento conectadas en intereses y formatos a los lenguajes documentales, en mayor medida que con anterioridad.

5. Conclusiones

Internet puede cambiar la concepción, los principios y los procedimientos del Análisis Documental y de los lenguajes documentales existentes hasta el momento. En efecto, crecen los niveles de estudio (el tipo de información, los niveles de lectura, la diversidad de situaciones documentales posibles, las estructuras, etc.) y se modifican los factores que concurren en este dominio (la nueva definición del usuario, el modelo de recuperación, el formato de documento, el propio discurso periodístico, el documentalista y los fondos documentales).

Asimismo, todos los datos parecen incidir en la conveniencia de mantener el método de empírico de trabajo ayudado por el recorte epistemológico sobre las disciplinas que conforman el cruce interdisciplinar que es, al fin y al cabo, la Organización y Representación del Conocimiento.

En general, se considera imprescindible la aplicación de desarrollos automáticos donde el factor humano estará presente en tres aspectos: a) en la interactividad existente entre el usuario, el productor y documentalista, b) en la aportación al trabajo documental, más en concreto al Análisis Documental y a los lenguajes documentales, puesto que el documentalista sigue cumpliendo una labor esencial y c) su utilización como modelo para la mejora de los procesos implicados en esta cuestión y, en definitiva, en el rendimiento de los sistemas documentales.

6. Referencias

Aramburu Cabo, M^a José ; Berlanga Llavori, Rafael (1997). An approach to a digital library of newspapers. // *Information Processing & Management*. 33: 5 (1997) 645-661.

- Codina, Lluís (1995). Recuperación de Información e hipertextos: sus bases lógicas y su aplicación a la Documentación Periodística. // Fuentes; M^a Eulàlia (edit.). Manual de Documentación Periodística. Madrid : Síntesis, 1995, p. 213-230.
- Codina, Lluís (1996). El Llibre digital: una exploració sobre la informació electrònica i el futur de l'edició. Barcelona : Centre d'Investigació de la Comunicació, 1996.
- Codina, Lluís (1997a). Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I). // Information World en Español. 6 : 5 (mayo 1997) 22-27.
- Codina, Lluís (1997b). Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte II). 6 : 6 (junio 1997) 18-27.
- Cornellá, Alfons (1996). Medios de Comunicación y empresa en la era de Internet. // Sánchez Casabón, Ana Isabel; Ubieta Artur, Isabel (eds.) Jornadas sobre Documentación y medios de Comunicación (actas): Zaragoza 20, 21 y 22 de mayo de 1996. Zaragoza : Universidad, 1996. p. 65-71.
- Dong, Xiaoying ; Su, L. T. (1997). Search Engines on the World Wide Web and Information Retrieval from the Internet: a review and evaluation. // Online & CDROM Review. 21 : 2 (1997) 67-81.
- Fulton, Katherine (1996). A tour of our uncertain future: <http://www.journalism.now>. // Columbia Journalism Review. 34 : 6 (march-april 1996) 19-26.
- Forrester, Michael (1995). Indexing in hypertext environments: the role of user models. // The Indexer. 19 : 4 (october 1995) 249-256.
- Fuentes i Pujol, M^a Eulàlia (1996). Posibilidades documentales de la red Internet en el ámbito periodístico. // Actas de las 5^a Jornadas Españolas de Documentación Automatizada, vol 2. Cáceres : Fesabid, 1996. p. 557-586.
- García Gutiérrez, Antonio (1992). Análisis documental del discurso periodístico. Madrid: CTD, 1992.
- García Gutiérrez, Antonio (1996). Procedimientos de Análisis Documental Automático: Estudio de caso. Sevilla : Instituto Andaluz del Patrimonio Histórico, 1996.
- García Jiménez, Antonio (1995). Metodología de Validación del Análisis Documental y de los lenguajes documentales en el discurso periodístico. Madrid : UCM, 1995. (Tesis doctoral).
- García Marco, Fco. Javier (1994). Hipertexto y lenguajes documentales: retos y sinergias. // Actas de las IV Jornadas Españolas de Documentación Automatizada : Gijón 6, 7 y 8 de octubre de 1994. Oviedo : Universidad, 1994. p. 417-426.
- García Marco, Fco. Javier (1995). Paradigmas científicos en representación y recuperación de la información. // García Marco, Fco. Javier (ed.). Organización del Conocimiento en Sistemas de Información y Documentación. Actas del I Encuentro de Isko-España: Madrid, 4 y 5 de noviembre de 1993. Zaragoza : ISKO-Es, 1995.
- García Marco, Fco. Javier; Tramullas Saz, Jesús (1996). World Wide Web: fundamentos, navegación y lenguajes en la red mundial de información. Madrid : RA-MA, 1996.
- Moya Anegón, Félix. Evaluación de buscadores ¿Cuál es el mejor?.
<URL=Http://www.ugr.es/-felix/g9/evaspa.htm>

- Nicholas, David ; Martin, Helen. Assessing information needs: a case study of journalist. // *Aslib Proceedings* 49 : 2 (February 1997) 43-52.
- O' Kane, Kevin C. (1996). *World Wide Web-Based Information Storage and Retrieval*. // *Online & CDROM Review*. 20 : 1 (1996) 11-20.
- Pinto, María ; Gálvez Carmen (1996). *Análisis Documental de Contenido: el procesamiento de la información*. Madrid : Síntesis, 1996.
- Scarela, Janne et al. (1997). Logical structure of a hypermedia newspaper. // *Information Processing & Management*. 33 : 5 (1997) 599-614.
- Schamber, Linda (1996). What Is a Document?: Rethinking the Concept in Uneasy Times. // *Journal of the American Society for Information Science*. 47 : 9 (1996) 669-671.
- Tramullas Saz, Jesús (1996). Periodismo e Internet: tipología, creación, características y perspectivas de los recursos de Información Periodística. // Sánchez Casabón, Ana Isabel; Ubieto Artur, Isabel (eds.). *Jornadas sobre Documentación y Medios de Comunicación (actas)* : Zaragoza 20, 21 y 22 de mayo de 1996. Zaragoza : Universidad, 1996. p. 87-104.
- Van Dijk, Teun A. (1990). *La noticia como discurso: comprensión, estructura y producción de la información*. Barcelona : Paidós, 1990.