



HAL
open science

Documento: forma, signo y medio, re-formulaciones de lo digital

Rogelio T. Pedauque

► **To cite this version:**

Rogelio T. Pedauque. Documento: forma, signo y medio, re-formulaciones de lo digital. 2003. sic_00001160

HAL Id: sic_00001160

https://archivesic.ccsd.cnrs.fr/sic_00001160

Submitted on 6 Dec 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Documento: forma, signo y medio, re-formulaciones de lo digital

Rogelio T. Pedauque. STIC-CNRS
Contacto : pedauque@enssib.fr

Version española, Regina Carreira (carreira@enssib.fr), 10-11-2004

Resumen

Este texto es el fruto de un trabajo colectivo de reflexión que se lleva a cabo en el marco de la red temática pluridisciplinar 33 del departamento STIC del CNRS. Se propone precisar la noción de documento en su tránsito al medio digital, a partir de un conjunto de investigaciones que privilegian, ante todo, la forma (como objeto material o inmaterial), el signo (como vehículo de sentido) o el medio (como vector de comunicación).

Cada uno de estos componentes refleja las grandes transformaciones que convergen en la actualidad y su superposición resalta la pertinencia de un enfoque pluridisciplinar para un análisis lúcido y completo de la noción de documento y de su evolución.

Contexto

Muy pocos artículos científicos proponen una definición actual del documento, y todavía es menor el número de quienes la discuten. El documento se identifica directamente como objeto de análisis en pocas comunidades científicas: la de los investigadores en ciencias de la información (o *Information Science*¹), al integrar trabajos que reflejan la amplia renovación en las técnicas documentales, como resultado de los procesos informáticos; la de los investigadores que se interesan en la digitalización y en los temas ligados a la indexación y a la categorización, extendiendo a menudo sus reflexiones a la gestión electrónica de documentos; y también la de quienes desenvuelven herramientas de edición electrónica (*Electronic Publishing*²). Por otro lado, y puesto que resulta un elemento indispensable para la construcción y la evolución de la disciplina, la noción de documento se discute en foros de historia y especialmente en arqueología o en geografía, para el estudio de cartas, e incluso en derecho, agrupando textos, artículos legislativos, reglamentos o circulares, guardando eso sí un enfoque estrictamente instrumental y raras veces directo.

Numerosos diccionarios, repertorios de normas y enciclopedias plantean definiciones, más próximas de la designación y de la descripción, que de la reflexión profunda de la noción.

Desde el *documentum* latino, vinculado etimológicamente a la enseñanza (*docere* = enseñar) hasta su acepción más reciente y marginal, la más frecuente, pero también imprecisa, la de “información”, todo parece indicar que la noción de documento se apoya en dos funciones: la prueba (el denominado “documento probatorio” de los juristas o clave de un informe) y la información (que se refiere a la representación del mundo o al testimonio). La archivística contemporánea, por ejemplo, le reconoce ambas funciones, admitiendo para el documento un “valor de evidencia” (de actividad) en un sentido más amplio que el de prueba jurídica, y un “valor de información”, que se corresponde con su acepción más general, arriba indicada.

Otros artículos científicos emplean un vocabulario diferente para designar objetos comparables, a veces definiendo el documento de forma rigurosa, aunque a menudo víctima de interpretaciones diversas. Así, los investigadores en informática, desde el estudio de redes, de bases de datos, de búsqueda de información y de tratamiento lingüístico hasta la ingeniería de conocimiento, al igual que otras comunidades como la de lingüistas, semiólogos y psicólogos

¹ Se mantiene la cursiva del texto original así como la terminología en francés (Ciencias de la Información)

² Trad. Inglés: “Publicación electrónica”

del aprendizaje; sociólogos de la cultura y de la organización, economistas de los medios o de la información, juristas de la propiedad intelectual y todos los “humanistas” en su conjunto, emplean una diversidad de palabras como información, dato, fuente, fichero, escrito, texto, imagen, papel, artículo, obra, periódico, hoja, página, etc; que, por supuesto, no son sinónimas y, por lo mismo, se justifican en el contexto particular de una investigación en curso, pero guardando siempre (aunque no se asuma en todos los casos) una relación con la noción de documento.

Por último cabe señalar la omnipresencia de los documentos en el día a día (especialmente en el plano administrativo y en la actividad científica). Se trata, así, de una noción intuitiva, no siendo necesario precisarla.

Tal vacío constituye hoy un problema. En efecto, la digitalización transforma profundamente la noción de documento, sin que podamos medir de manera rotunda sus efectos y consecuencias; un defecto por haberse precipitado en la delimitación de la problemática. Del papel, el soporte más corriente, al documento digital, estas transformaciones se localizan fácilmente; por su aspecto material, por ejemplo, o por el tratamiento cognitivo, la percepción e incluso el uso que se hace de él. Aunque anunciada en los textos de algunos pioneros y preparada para una convergencia cada vez mayor entre lo escrito y lo audiovisual, esta cuestión todavía es reciente, para cuanto más caótica y sin que ya sea posible dar marcha atrás. Es probable que los numerosos investigadores que abordan estas cuestiones bajo múltiples facetas se beneficien de una visión de grupo, permitiéndoles así posicionarse de forma más lúcida.

El retraso de este análisis puede explicarse por el contraste entre una relativa estabilidad, dominante hasta hace poco, en contraste con la velocidad y la radicalización de los cambios actuales. Entonces, sólo cabía la posibilidad de interrogarse como historiador sobre un objeto demasiado corriente como para resultar complejo; actualmente, aún no hemos sido capaces de analizarlo guardando las distancias.

El documento se ha constituido como un objeto cuya materialización más banal es la hoja de papel, en el transcurso de un proceso secular en el que útiles, saberes y estatutos se van entrelazando. Con el tránsito al medio digital, desde hace ya algunas décadas, hemos entrado en una nueva fase, en la que determinadas características siguen en dependencia directa con el periodo anterior mientras que otras, al contrario, anuncian un cambio radical y quizás la emergencia de una noción diferente, que retome en totalidad o en parte, la utilidad social de lo que llamamos “documento”. La manifestación más simple de este cambio es la pérdida de la estabilidad del documento como objeto material y su transformación por medio de un proceso de construcción a la demanda, que cuestiona por momentos la confianza que se le otorgó al inicio.

La cuestión entre ruptura y continuidad no se plantea sólo en relación al objeto. Los métodos de análisis y los epistemológicos están también en rápida evolución.

Una reflexión pluridisciplinar

Desde nuestro punto de vista, estas dificultades sólo pueden articularse mediante una reflexión decididamente pluridisciplinar. El departamento STIC del CNRS nos reconforta en esta línea, al lanzar una red temática pluridisciplinar, llamada “Documentos y contenidos: creación, indexación, navegación” (<http://rtp-doc.enssib.fr>) que agrupa a un centenar de investigadores. Para algunas de las disciplinas que lo integran, el documento no constituye la noción central y los investigadores sólo tienen una relativa aprehensión. El objetivo de la red es, así, el de desplazar esta atención oblicua para hacer del documento un tema principal de investigación, al menos durante un tiempo, intercambiando las aportaciones parciales de unos y otros.

No es seguro que exista, entre disciplinas o incluso en el seno de cada una de ellas, un consenso relativo a las cuestiones que abordamos. Nuestro objetivo no se basa en la uniformización, la definición de una línea, una corriente o una escuela de pensamiento, sino en la aclaración y en la precisión de los conceptos, con la intención de establecer los malentendidos, de abrir nuevas perspectivas y de identificar los posibles desacuerdos. Nuestra convicción es que el diálogo interdisciplinar sólo resulta fructífero si se logran despejar las nociones esenciales, para discutir las o para utilizarlas como punto de apoyo.

Semejante tentativa resulta arriesgada. Por un lado, los contrasentidos o simplemente la superficialidad están al orden del día. Por el otro, las diferentes bases de las disciplinas o de las corrientes pueden resultar contradictorias entre sí. A las dificultades conceptuales pueden añadirse obstáculos de carácter práctico. Cada especialidad desarrolla de forma natural su propia cultura y su propio vocabulario, por buenas (el rigor) o malas (la protección) razones. Las mismas palabras se revisten por momentos de sentidos diferentes en función de las comunidades y a menudo resultan extranjeras para quienes resultan ajenos a ella. En el marco de un texto transversal, nos limitamos al empleo de un vocabulario común, en el sentido más amplio del término, con el riesgo de poder caer en la caricatura.

Concretamente, este texto es el resultado de un trabajo colectivo en el seno de la red. Dada la metodología empleada para su redacción y las numerosas contribuciones que lo integran, hemos optado por omitir toda cita o referencia bibliográfica directa. De hacerlo, la dinámica colectiva se desviaría, induciéndose una concurrencia entre autores o escuelas de pensamiento. No obstante, se propone una bibliografía desde el sitio del rtp-doc.

Propuestas

Organizaremos nuestras propuestas por analogía, con la distinción que se hace en lingüística entre sintaxis, semántica y pragmática. Sin entrar en un debate sobre la validez o no de tal paralelismo, o sobre la legitimidad de esta clasificación en ciencias del lenguaje, constatamos que esta división simplifica la organización de las investigaciones actuales, al igual que la de las corrientes que las fundamentan. Distinguiremos:

- El documento como forma; en esta categoría se agrupan aquellos enfoques que analizan el documento como objeto, material o inmaterial, y que lo estudian en función de su estructura con el fin de mejorar su análisis, uso y manipulación.
- El documento como signo; estos investigadores valoran primero el documento como vehículo de sentido y como expresión de una intencionalidad. Así, el documento y el contexto del sujeto, que lo construye o reconstruye aportándole nuevos sentidos, resultan indisociables. Paralelamente, el análisis se realiza en el marco de un sistema documental o sistema de conocimientos.
- El documento como medio. Por fin esta dimensión plantea la incógnita sobre el estatuto del documento en las relaciones sociales. El documento es la huella, construida o hallada, de una comunicación más allá del espacio y del tiempo. Es, al mismo tiempo, un elemento de los sistemas identitarios y constituye un vector de poder.

La analogía con la lingüística todavía es vaga. Podría argüirse que la primera categoría se refiere a la morfosintaxis y la segunda, a la semántica y a la pragmática al mismo tiempo. No sirviendo como fundamento, nos basta con que la comparación resulte eficaz en el terreno.

Cada categoría debería apreciarse como una tendencia dominante y no como una dimensión exclusiva. Así, los investigadores que tratan el documento a partir de la primera óptica no subestiman necesariamente las otras dos. No obstante, su análisis y razonamiento la privilegian, resultando las otras complementarias y apremiantes.

Precisamente, el término más apropiado sería el de “entrada”³. Cada una de ellas constituiría en efecto una forma de tratar el objeto de la investigación, el documento, en el que el resto de las disciplinas se encontrarían con motivo de planteamientos, coacciones, obstáculos o repentinas barreras en contra de la primera reflexión. Pero la tendencia de cada enfoque consiste también en relativizar las otras.

Cada categoría se abordará en función de un mismo esquema:

- En primer lugar, la identificación de las principales disciplinas, conocimientos o especialidades que privilegien este punto de vista. No se trata aquí de discutir su

³ Se ha optado por guardar la equivalencia literal del término francés “entrée” que se identificaría con la noción de rúbrica española, en el sentido de categoría que sirve a la clasificación

validez o su estatuto científico, sino de examinar la diversidad de trabajos representativos en el tema, sin juzgar su importancia o valor.

- Después, plantearemos una interpretación de la evolución entre los puntos de vista, en el tránsito del documento tradicional al documento digital.
- A partir de cada entrada construiremos progresivamente una definición del documento.
- Resaltaremos algunos de los temas por resolver dentro de cada categoría, más allá de la indagación de las investigaciones en curso.

A cada etapa intentaremos despejar lo esencial, sin detenernos en pequeños matices, excepciones o casos particulares. El objetivo consiste en resaltar los que son fundamentales y no ser exhaustivos. Por lo que respecta a la definición, uno de los métodos se basaría en una búsqueda sistemática de los casos que no le corresponden para proceder, así, a una definición universal. Este modelo no parece ser el más operativo. Nuestro objetivo no pretende responder a todas las cuestiones, sino construir una definición genérica, aun con el riesgo de llegar sólo a identificar las excepciones de casos particulares, situaciones intermediarias, o simplemente los análisis incompletos y erróneos.

Como conclusión, propondremos una síntesis de tres entradas que tiene como meta destacar los elementos de continuidad y de ruptura en relación a la época precedente.

Documento como forma

El término “forma” puede revocarse y llama a confusión. Lo utilizamos al no haber encontrado otro que sea más pertinente. Debe entenderse aquí como “contorno”, incluso en el sentido de “figura”. Dicho de otro modo, la forma se aprehende como objeto o como una inscripción dentro de él, dentro del que se identifican otras, en relación a un “formalismo”; es decir, que este objeto o inscripción responde a las reglas que lo constituyen.

El documento se percibe aquí como un objeto de comunicación que se rige por reglas de formalización más o menos explícitas, materializado por un contrato de lectura entre un productor y un lector. El documento se estudia en el marco de un protocolo implícito de comunicación, con independencia de su contenido textual o contextual.

Relación de especialidades

Conviene subrayar desde el principio el lugar específico de la escritura, técnica cuyo aprendizaje compartido emplaza el documento, desde su aparición, en un contexto social preferente.

Resultan numerosas las prácticas, profesionales o no, que privilegian este punto de vista; algunas tan antiguas como la caligrafía, la tipografía; otras formas de representación, como las técnicas musicales, audiovisuales, cinematográficas, o la biblioteconomía, que tiene como eje motor el catálogo, la clasificación y la gestión documental e incluso la diplomacia archivística.

Resulta lógico el que los informáticos, que parten de objetos materiales para digitalizarlos, los referidos a la imagen, guarden fuertes afinidades con las primeras especialidades. Se interesarán enseguida por la estructura del documento en el marco de los sistemas automáticos de identificación de la forma. En primer lugar, localización automática de caracteres; luego, la caligrafía, llegando así al tratamiento formal de imágenes. En esta disciplina se enfrentarán a problemas relativos al formato, al intercambio, al almacenamiento, a la descripción, al destino, a la conservación o al tratamiento masivo. Se trata pues de la lectura automática o del análisis de documentos. Los investigadores intentan decodificar el objeto, especificando y explotando en paralelo el protocolo de comunicación subyacente (que es el contrato de lectura).

Del mismo modo actualizarán sus prácticas los que se interesan en la compaginación, en las formas editoriales y en la normalización internacional de estas áreas; también los interesados en el tratamiento de texto y quienes articulan sistemas digitales de vídeo.

Otras áreas de especialización informática han optado por este primer enfoque. La concepción de los sistemas electrónicos de gestión documental, como indica su nombre, nace de la idea de que el documento preexiste como ente identificable, aun pudiendo ser

virtual. Aunque esta vez el referente ya no es un objeto concreto, sino un fichero informático, un gran número de los problemas que aquí se plantean confirman las mismas incógnitas fundamentales.

Por último, asistimos a un cambio a gran escala desde la aparición del World Wide Web. En torno a este tejido se han ido desarrollando intensas actividades de búsqueda, concepción, negociación, normalización y desarrollo; con especial y no único interés, en el seno del consorcio WC3. Aunque los investigadores emplean poco el término de “documento”, perfiriendo el de “fuente”, que aglutina otros muchos objetos, gran número de los interrogantes que en su día se plantearon los conceptores del Web, en su versión actual (es decir, antes del llamado “Web semántico”), se erigen también fruto de este primer enfoque. Se trata de averiguar cómo asociar entre sí, y a escala planetaria, las diferentes fuentes para así definir las normas y los sistemas aplicables a todas las máquinas, dando una dirección que sea identificable a dichas fuentes. De éstas, muchas conservan las características del documento, tal y como se han entendido en esta primera dimensión: archivo HTML o XML, imágenes, grabaciones sonoras o de vídeo, “flujo”⁴ multimedia, etc.

Evolución

Una primera definición del documento podría traducirse por la ecuación: *Documento tradicional = soporte + inscripción*. En efecto, el acento recae primero en el soporte manipulable (en sentido puro) al que se ha fijado una huella que se interpreta formalmente por un ojo, un oído humano, o en el curso de una lectura braille. Otros sentidos lo harán, por qué no mañana, con o sin prótesis. La huella representa el contenido, fruto de tal inscripción.

El soporte tradicional dominante (que no exclusivo) es el papel, la traza, la escritura, manuscrita o impresa. La hoja escrita, el elemento de base, se enriquecerá gracias a la compaginación y al paratexto, para de ahí extenderse al libro, a la distribución, etc., confiriendo gran plasticidad y complejidad al documento. El “codex” (o libro encuadernado) constituye sin ninguna duda la forma más elaborada del documento tradicional. Su calidad sería proporcional al peso de sus “particularidades” ¡Sin mayores cambios desde hace casi un milenio!

Gracias al importante esfuerzo social (la escuela, que permite la adquisición de la lectura) este documento se percibe de manera directa; es decir, sin que intervengan herramientas sofisticadas (a excepción de las gafas para algunos) para una muestra más o menos representativa de una sociedad dada: la que ha aprendido a leer.

En la historia, cuando esta noción (*soporte + inscripción*) se extendió a otras formas de representación, como la reproducción musical, el cine y luego el audiovisual, el soporte no guardó su facultad de apropiación directa. Aun cuando la representación se aproximaba a la percepción humana inmediata (necesitándose un aprendizaje más ligero para decodificarlo), el dispositivo de lectura se sofisticó. Es indispensable el disponer de máquinas para escuchar un disco (reproducido digitalmente), proyectar una película (fijada en el negativo) o una cinta de vídeo (grabada⁵). El objeto sigue siendo necesario en la lectura, pero ya no es por sí suficiente.

Mejor aún, la radiodifusión y luego la televisión permitieron separar la codificación de la transmisión de la señal. Así, oyente y telespectador escuchan y ven un programa cuya emisión les escapa. No es dueño y señor de su lectura, salvo puntualmente, a causa de la grabación en banda magnética. En cierta manera, la difusión hertziana, al entrar en su domicilio, pierde una parte de la autonomía espacio-temporal que había ganado, al haberse manipulado los objetos, digital o analógicamente.

⁴ En un sentido abstracto o figurado el término “flot” francés equivale a desbordamiento, torrente. En este caso ha sido traducido por *flujo continuo* de datos intercambiados en el Web (R.A.L.).

⁵ El castellano no conoce la misma diferencia que el francés entre *graver* (soporte digital, discos ópticos), *enregistrer* (en banda magnética, de la radio o del magnetoscopio). Para paliar tal diferencia se ha optado por diferenciar los términos en función del tipo de soporte: reproducción (digital), grabar (analógico).

El audiovisual ha abierto así la vía a una evolución en la utilización de los soportes, pero la mutación esencial para nosotros se constituye con el paso de la inscripción de una señal analógica a la de una señal digital, consecuencias radicales para el conjunto de documentos, escritos, imágenes y otros productos audiovisuales. Estas mutaciones se localizan tanto en los dispositivos de lectura y escritura, como en los propios documentos.

En relación a los dispositivos, se observa en primer lugar un cambio recíproco de lugar entre lo escrito y lo audiovisual. El primero incluye un dispositivo que le resulta familiar al segundo. Ya no es posible leer sin máquina. Aunque la producción impresa pase por un fuerte equipamiento técnico, la lectura en papel, ya lo hemos dicho, es directa, o casi. Discos ópticos o magnéticos, bandas registradas, máquinas que tratan y restituyen la señal, así como las conexiones en red, constituyen los utensilios imprescindibles para adquirir una lectura electrónica individual, dispuesta a recuperar su estado anterior gracias a la impresora. Oyente y espectador pueden, por su parte, conjugar en el Web, el dominio del lazamiento y detención de un flujo temporal perdido en la red de radiodifusión, que transmite los programas de forma ininterrumpida, hasta ahora sólo modulable mediante la grabación vídeo.

La segunda consecuencia de los dispositivos es el cruce entre soportes y señales. La noción de soporte se hace compleja y se vuelve ambigua ¿Es el fichero, el útil material que lo alberga, o se trata más bien de la superficie de la pantalla que lo exhibe? Al atravesar la red, el documento se copia por fragmentos en los conmutadores con una duración variable. Los mismos “soportes” pueden además albergar sin diferencia toda representación, a condición de que ésta sea digital e incluso con reservas de compatibilidad de formato. Las representaciones, también ellas, pueden confundirse: podrán “leerse” así, y en estrecha relación, texto, imagen, audio y animación.

Mientras que antaño la imprenta había privilegiado el soporte material, dado el peso tecnológico que es inherente a toda actividad productora de documentos, la publicación electrónica ha permitido la aparición de una producción a la carta (en pantalla o en papel indiferentemente). De ahí que el soporte papel haya perdido la primera posición en beneficio de la publicación electrónica. En esta línea, se puede retomar el hecho de que una de las evoluciones importantes de la publicación asistida por ordenador haya sido la irrupción del *wysiwyg* (*What you see is what you get*⁶) que ha hecho posible la visualización en la pantalla como en el papel.

Por último, desde el desarrollo conjunto de la microinformática y de las telecomunicaciones, las máquinas, también ellas, se multiplican y se vuelven autónomas; ordenadores portátiles, hojas de cálculo electrónicas, teléfonos y utensilios integrados de diversos tipos están en búsqueda de una mejor adecuación entre el comportamiento de los lectores y de sus necesidades, generales o específicas. En la evolución de los documentos resulta imprescindible el que la difusión del teléfono móvil haya sido mucho más rápida y de mayor alcance que la del micrordenador.

La noción de soporte ha perdido, así, su nitidez primera. Pero en el marco de nuestra ecuación (*soporte = inscripción*) las consecuencias que lo digital tiene para el segundo término, el de inscripción, son del todo radicales. La inscripción implica codificación, operación familiar en la comunidad informática. Los informáticos están en la obligación de aislar los elementos lógicos que constituyen esta dimensión del documento para darles forma, automatizar sus operaciones y combinar de nuevo los que son diferentes, de este modo perfeccionados.

En relación a este capítulo, podríamos extender un vínculo con la noción de “programa”, tal y como se presenta habitualmente en informática: *programa = software + datos*. Un documento sólo sería un caso particular de programa, en el que la parte “software” representaría la “estructura” y la parte “datos”, el contenido. La ecuación daría lugar: *documento digital = estructura + datos*. Coherentes con esta primera entrada sobre la forma, los investigadores no intervendrían sobre el contenido; al contrario, estudiarían la estructura, por definición susceptible de ser modelada, y que representa en alguna medida,

⁶ Trad. Inglés: “Se toma lo que se ve”

con independencia del soporte, el contrato de lectura que se establece entre el productor del documento y sus lectores potenciales.

La estructura varía enormemente en función del tipo de documento. Algunos están poco o nada estructurados; por ejemplo, ciertas obras de arte o textos espontáneos en los que fondo y forma resultan indisociables; otros, al contrario, siguen un formalismo estricto. La estructura también difiere según el tipo de medio. El audiovisual introduce, por ejemplo, una dimensión temporal que apenas se manifiesta en el medio escrito. No obstante, el análisis ha permitido localizar y aislar varios niveles de estructuración en casos generales. Éstos se han construido a partir de dos líneas de investigación; los primeros parten de lo analógico para llegar a lo digital, los otros evolucionan en sentido inverso. Antes de retomar la noción de estructura, conviene entender la lógica de su razonamiento. A la primera corriente se le ha atribuido como tarea el traducir a un formato digital los documentos tradicionales con el fin de aprovechar sus prestaciones técnicas. Dicho de otro modo, un documento tradicional tendría que evolucionar de una ecuación a otra: de *soporte + incipción* hacia *estructura + datos*. Esta operación puede hacerse a partir del primer documento que se digitaliza. Se intenta precisamente desmaterializar el documento, apoyándose en una táctica que se basa en el tratamiento de imágenes y en el reconocimiento de formas. Cabe también aquí una reflexión a partir simplemente de la representación del documento, que construye de manera directa, y sin ayuda de los soportes anteriores, el equivalente visual de su representación, total o parcialmente. Anótese que esta operación no es trivial desde el punto de vista social. Y que ha de efectuarse en los dos sentidos, fundamentalmente por razones de índole jurídica que retomaremos en el transcurso de la tercera entrada. En esta primera corriente encontramos a los técnicos de la imagen que, como su nombre indica, centran su investigación en la reconstrucción de la imagen del documento y, por ende, en su representación formal

El principio es el reconocimiento de formas. Para ser reconocida, una forma debe conocerse de manera previa. Cuanto más se apoya el documento original en estructuras que son genéricas, más fácilmente se podrá transponer. Así, la complejidad se acrecienta al pasar de caracteres tipográficos a gráficos, de ahí a los esquemas, después a la imagen y por fin a los objetos en tres dimensiones. Aunque el objetivo sea el de reproducir una percepción similar u homóloga a la del objeto original, se trata no obstante de una nueva traducción que podrá ocultar elementos significativos o, al contrario, dar a conocer y redescubrir otros nuevos, conforme a las elecciones tecnológicas que se han efectuado y al uso posterior de los ficheros.

Otros investigadores parten directamente de la ecuación final (*documento digital = estructura + datos*). Dicho de otro modo, emprenden el camino inverso. A partir del epicentro de la reflexión informática, que es la elaboración del algoritmo, reconstruyen los documentos, retrocediendo paso a paso en su lógica o estructura interna, para llegar así a una representación que resulte legible en la pantalla. Esta segunda corriente forma parte del uso común que los lenguajes informáticos hacen del texto, al integrar progresivamente la preocupación formal (*wysiwyg*). Dicha corriente dio lugar primero a la burótica; entró en contacto con los investigadores que desarrollaban las herramientas para la publicación electrónica y, por fin, se ha confrontado a la necesidad de poder intercambiar los documentos a gran escala. Así es cómo verdaderamente estalla, con la revolución del Web.

Estos informáticos se sirvieron de un razonamiento por estratos o capas, con el propósito de aislar y tratar por separado los elementos de la estructura del documento. Han descubierto o redescubierto los diferentes niveles lógicos de esta última, siendo el nivel inferior el del texto o señal analógica, que se ha intentado unificar bajo la forma de código unitario, MPEG, etc.

La noción de balizaje del contenido, por su parte, se vincula a la estructura del documento, en el momento en que cada componente fotográfico de carácter electromagnético se transforma en fotocomponente de tipo digital. Ha instaurado poco a poco dos principios: las balizas describen mejor la estructura que las características físicas del documento y se comprenden tanto por el programa como por el intérprete humano.

Sin trazar de nuevo esta evolución, digamos que desde el punto de vista que aquí nos ocupa, el Web se puede describir como una infinidad de documentos asociados entre sí. Su arquitectura se apoya en tres pilares: las fuentes identificadas por un esquema de destinación universal (identificación), que se representa por un conjunto no exclusivo de esquemas (representación) y que se intercambia a partir de protocolos estandarizados (interacción). Esta arquitectura supone que se puede acceder a los documentos desde todas partes y sobre cualquier tipo de material, en función de los grupos de usuarios y de sus particularidades.

Ambas corrientes, la del reconocimiento de documentos tradicionales y la que pretende la construcción directa de documentos digitales, son dependientes y convergen en un mismo fin, aun partiendo de puntos diferentes. Han permitido resaltar dos niveles fundamentales en la estructuración de documentos: la estructura lógica (la construcción de documentos en fragmentos o subfragmentos articulados entre sí) y la representación formal de la presentación, los “estilos”, en sentido informático; en el texto, por ejemplo, las elecciones tipográficas. En nuestra meta, la revolución fundamental quizás sea la uniformización progresiva del formato del documento (en el sentido informático) que permite, justamente, un tratamiento simple a ambos niveles.

Un documento tendría que ser fácil de leer sobre cualquier ordenador y descifrable por las distintas aplicaciones. La tendencia es la de la fragmentación: los formatos “propietarios” invaden el mercado, limitando los que son universales y gratuitos a los especialistas. Por otra parte, los formatos “no universales” conllevan a situaciones de ilegibilidad: un programa no puede leer un fichero, una aplicación no consigue abrir un documento, una página Web no se pega correctamente sobre la pantalla. Añádase a esto que el formato ha de transcribir el alfabeto: el “formato” se escribiría en varias lenguas. La normalización es esencial.

Es posible que el éxito de la norma XML, cada vez más generalizada, así como el de sus numerosos derivados, marque no sólo una nueva etapa, sino el resultado de estos movimientos. En efecto, esta norma, que forma parte a la vez de la informatización de las técnicas editoriales (SGML) y de la sofisticación del primer balizaje en el Web (HTML) integra en un mismo fichero estructura y contenido, mediante un balizaje normalizado del texto, permitiendo encontrar y superar ampliamente la plasticidad y la complejidad de la encuadernación, señalada al comenzar esta parte y de la que hemos omitido algunas funcionalidades a lo largo del discurso. Pero asimismo, introduce cuestiones inéditas al reformular los términos del antiguo contrato de lectura en el que se había perpetuado el vínculo entre la representación percibida y la estructura lógica, por medio del soporte. Con un enfoque XML, se capturan estructura y contenido; la forma puede deducirse de otras diferentes. No se representa de manera intrínseca. Se dejaría a un lado así la forma como elemento primordial del documento. Pero, por otro lado, muchos trabajos se centran en los diferentes medios de representación y producción de la forma del documento electrónico, sobre todo cuando se trata de documentos XML.

Así, nuestra ecuación inicial se transformaría de nuevo: *documento digital = estructura + datos* daría lugar a *documento XML = datos estructurados + compaginación*, en la que el segundo componente (el “estilo”) podría modularse en mayor medida. En el marco de XML, forma y estructura se definen por separado, teniendo como intermediaria a la hoja de estilo (XSL).

Una evolución posible, aunque no la única, consistiría en que los documentos así redactados se integrasen en una base de datos centralizada o distribuida, y que el conjunto de ficheros se apropiase de uno o varios juegos “lego” en el que las diferentes piezas, de tamaños, formas y usos distintos, se corresponderían entre sí, y en función de varias configuraciones. Estaríamos a punto de dar el gran paso adelante, pues un documento sólo tendría una forma propia en dos momentos: el de la concepción por su autor, que ha de idearlo y comprenderlo para asegurarse de que se corresponda con sus elecciones (no siendo ni siquiera necesario, en caso de que el documento sea el resultado de un proceso) y el de reconstrucción por el lector. Es poco probable que el documento resulte idéntico en ambos casos. Otra evolución posible consideraría al documento como la base de datos en el que las diferentes entradas sólo representan una parte de su riqueza. Una comunidad de investigadores estudia este

tema, en el contexto del Web semántico, bajo la terminología de “documentos virtuales, susceptibles de ser personalizados”.

Esta evolución plantea la problemática de gestión del tiempo para uno o varios tipos de documentos, así como la de su escritura, riqueza e incluso re-escritura⁷ por diferentes sujetos. La gestión de las versiones sucesivas de un documento es ya de por sí delicada para las personas, organizaciones o en el marco del Web. Se trata aquí de idear los procedimientos que permiten asociar un texto a un autor (o a un colectivo de autores), permitiéndoles al mismo tiempo a cada uno apropiarse y reapropiarse de la totalidad o parte de los documentos que se producen en su seno, evitando así el “ruido”, fruto de las diferentes versiones sobre la misma información en la red. Así, la naturaleza y los orígenes de las modificaciones del documento actual se identificarán bajo una óptica coherente, con independencia de su formato y estatuto, ajeno a cualquier control centralizado.

Se perfilan los preceptos de esta última etapa y se intuyen también los problemas que de ella derivan. Nos aventuraríamos al prever su evolución y sus consecuencias, a no ser para señalar que éstas serán importantes y permanentes a corto plazo.

Definición 1

La observación de esta primera dimensión nos lleva a formular una definición del documento, en este estado aún incompleta, aunque representativa del importante cambio actual. Esta definición debe tener en cuenta la marginalidad del soporte; la misma función fundamental interpretada en el sentido inverso, a partir de la articulación entre estructura lógica y estilos, que definen así el contrato de lectura, entendido aquí como contrato de legibilidad.

Se llama documento digital al conjunto de datos organizados a partir de una estructura estable, asociada a reglas formales que permiten una mutua legibilidad entre su conceptor y sus lectores.

Esta definición es demasiado extensa para resultar al mismo tiempo nemotécnica. Recordemos las transformaciones de nuestra ecuación:

Documento tradicional = soporte + inscripción en Documento digital = estructura + datos.

Sugerimos la evolución actual, cuyas consecuencias son por el momento inciertas:

Documento digital = estructura + datos daría lugar a Documento XML = datos estructurados + forma. Recordemos que la norma XML, stricto sensu, no significa “forma”; ésta nace de la aplicación XSL.

Incógnitas

Este primer enfoque, por la forma, deja varias incógnitas por resolver. Insistiremos aquí en aquéllas que se refieren a la relación entre el mundo perceptible y la organización digital del nuevo medio documental.

Un primer grupo de cuestiones concierne la fijación de imágenes. Mientras que la “bibliografía material” ha estudiado de cerca el “objeto libro” en todas sus variaciones, todo parece indicar que en el tránsito al medio digital se haya profundizado en el tema de la estructura desde la rúbrica lógica, y a partir de ahí, en el tratamiento del documento. Así, en esta primera dimensión, los investigadores consideran con gusto que cualquier fijación es posible, dado que la estructura se ha integrado en el fichero. Los problemas de percepción constituirían otra problemática. Esta concepción, que roza la caricatura, propondría que estructura y contenido fuesen independientes, lo que como mínimo es discutible. La forma comprende sentido y algunos se interesan, desde hace tiempo, por ejemplo, en la importancia cognitiva de las posibles evoluciones en los llamados vínculos de navegación hipertextual.

⁷ Se mantiene la morfología del original francés (“re-écriture”) por razones de estilo y coherencia global del artículo

No obstante, queda por llevar a cabo un trabajo sobre la lectura electrónica para entender mejor los mecanismos de interdependencia entre los dos términos de la ecuación.

Cabe destacar que esta separación destruye las bases de la llamada diplomacia archivística, que tiene como fin la autenticación del fondo del documento a través de un análisis de su forma. Se deduce de ello que la autenticación (validación) debe (deberá) asegurarse por otros medios, de tipo técnico (filigrana electrónica) o de organización (archivistas intermediarios certificados) ¿Podría alguien algún día imaginar que las exigencias de la forma –entre ellas, la forma autenticada- se impongan y se validen por la existencia de una hoja de estilo “firmada” o en “filigrana”?

Estas cuestiones son tanto más delicadas por cuanto un mismo documento no puede leerse de forma corriente en distintos aparatos de lectura. Así, ¿ha de razonarse como si sus terminales no ejercieran algún tipo de influencia sobre la percepción? Basta comparar las pantallas de un microordenador, de una hoja de cálculo electrónica, de una agenda electrónica o de un móvil para comprobar lo contrario. Topamos, entonces, con aquellos soportes de lectura que creyéramos haber superado. Tales incógnitas son objeto de concurridos debates, especialmente, en el seno del consorcio W3C (*device independence*).

Los progresos obtenidos en la distribución en pantalla, a raíz sobre todo de los trabajos realizados por el laboratorio Parque Xerox, seguidos por el perfeccionamiento de los útiles de oficina, se deben a la permanencia de una organización visual; para el documento, se trata de la compaginación y de la encuadernación. Algunos técnicos gráficos proponen interesantes composiciones. Pero estos esfuerzos parecen estar muy poco vinculados a las investigaciones anteriores. Al igual, los libros electrónicos y todas las esperanzas puestas en el cartucho electrónico no han desembocado por el momento en aplicaciones irrefutables, aunque la existencia de un “codex electrónico” siga haciendo soñar a algunos. Aquellos para los que la representación espacial resulta fundamental ya se plantean el problema, como geógrafos y arquitectos, que son una excepción y no constituyen la regla general. Nuevamente, el audiovisual abriría prometedoras vías con el fenómeno de la realidad aumentada en el que se integran a la vez aspectos analógicos y de reconstitución digital.

Una segunda batería de cuestiones concierne la “vida permanente” de los documentos digitales. Estas incógnitas se debaten con frecuencia. Por un lado, no difieren de las problemáticas tradicionales sobre el archivaje y la conservación, simplemente se transponen a otras técnicas. Por otra parte, plantean temáticas del todo innovadoras: los ficheros XML, por poco que se actualicen y se conserven en un mínimo de condiciones, resultan teóricamente inalterables puesto que contienen la totalidad de sus informaciones en forma digital. Así, muchos consideran que los problemas de inalterabilidad se resolverían con muy pocos medios. Lejos se está de la idea en la que estos ficheros representen la (o las) forma(s) en la que se leen los documentos. Se podría deducir que la memoria completa de estos documentos consistiría en conservar la totalidad de los materiales y de los sistemas de lectura sucesivos que permiten acceder a ella. Aquí también queda por llevar a cabo un arduo trabajo teórico y práctico. Subrayaríamos por último un último conjunto de cuestiones. El documento tradicional es un objeto material manipulable. Éste desaparece con el medio digital, para convertirse, en su última fase, en una especie de rompecabezas en el que las piezas se distribuyen en función de las demandas del lector. No obstante, el lector tiene siempre acceso a partir de una máquina, o terminal en el que el documento se fija. ¿Acaso nos encaminamos hacia la versión extrema de esa idea que consiste en ver el documento como un avatar moderno o piedra milagrosa en el que los elementos significativos de tipo multimedia se fijarían, limitados únicamente por una lógica de sentidos y significados específicos? ¿O se trata de una reestructuración de documentos “tipo” que respondan a las necesidades o a las situaciones particulares en los que toda dinámica se verá confinada a funciones definidas? ¿Podría sostenerse la hipótesis según la cual la estabilidad visual del papel, la manejabilidad, así como la coexistencia de las hojas juegan una tarea fundamental en las actividades cognitivas? Por lo tanto ¿Acaso no deberían aunarse todos los esfuerzos con el fin de crear un “codex electrónico”? ¿Qué impacto tendrán en nuestros sistemas de conocimientos, las nuevas formas de lectura? ¿Qué hacer de la responsabilidad, jurídica o moral, del autor (individual o colectivo)? Y ya en relación

con nuestra rúbrica sobre la forma: ¿Es posible desvincular la elaboración de un documento de su forma perceptible? En consecuencia ¿Resultaría concebible la idea de ruptura formal entre la elaboración del autor (que también es el primer lector) y la proposición que se les hace a los lectores? El éxito de los formatos de tipo “facsimil” (PDF) se analiza a menudo como resistencia momentánea al cambio ¿Acaso no se trata aquí de una necesaria estabilidad perceptiva?

Estas incógnitas podrían resumirse en una única cuestión: Al descartar el soporte, ¿acaso no hemos dejado también a un lado la forma?

Documento como signo

Al igual que en la entrada precedente, el título de esta parte no debe tomarse en su acepción más académica. Desde hace mucho tiempo, el signo es objeto de análisis en trabajos científicos. Aun cuando los resultados de algunos de ellos nos sirven, no nos centraremos aquí en el concepto. Nuestro objetivo es el de reagrupar y el de presentar las investigaciones que consideran el documento como un objeto signficante.

La rúbrica que nos interesa aquí es la que refiere al tratamiento del contenido. Aunque la forma en ocasiones también se tenga en cuenta, sólo se hace por cuanto ella es portadora de sentido.

Especialidades relacionadas

Esta categoría engloba aquellas disciplinas sensiblemente diferentes de la primera, presentando algunas de ellas como de gran avance histórico en relación a la primera; como si, al pasar de la forma al signo, nos aproximáramos más al nudo de la problemática.

Así, en relación a las prácticas profesionales, la biblioteconomía da lugar a la documentación y luego a los profesionales de la información, más interesados en suministrar respuestas a los lectores que en gestionar los objetos mismos. La misma GED se convierte en *Knowledge Management* (KM)⁸ que, más allá de un sistema de conservación de ficheros, se trata de un proceso por el que los conocimientos útiles de organización se identifican de forma directa. Fundamentalmente, el Web adquiere un nuevo adjetivo que lo cualifica, el de “Web semántico”, dando a entender por ahí que un mejor uso de las capacidades de las máquinas en conexión, posibilitaría el tratamiento del contenido de los ficheros en línea, permitiendo una organización de los servicios, próximos a las demandas de los internautas.

Desde el punto de vista académico, esta categoría reúne primero a los especialistas que trabajan sobre el texto, la palabra, la imagen; en consecuencia: lingüistas y semiólogos de todas las tendencias que se especializan en el análisis del discurso, en la lingüística de “corpus”, en la semántica, así como aquéllos que se encargan de la configuración de herramientas para permitir un tratamiento automático de la lengua, dando lugar a la traducción y a los sistemas de búsqueda informáticos.

Otro vínculo se tiende en una segunda categoría de informáticos, encauzados en el área de investigación denominada “inteligencia artificial” que, a partir de la modelización del pensamiento, intentan construir herramientas capaces de dar respuesta por medio de la búsqueda directa en los ficheros. En paralelo, de la noción de información se pasa a la de conocimiento, con una ventaja sobre la primera, al integrar también el razonamiento. Una nueva disciplina emerge, la bautizada como “ingeniería del conocimiento”.

Pronto, la necesidad y la pertinencia de llevar a cabo una investigación de la información en la información resultará más urgente que nunca. Tanto el catálogo, que ha dado lugar a la indexación, como el tesoro, antecedente de las llamadas ontologías y de los “metadata”; son herramientas y objetos de búsqueda esenciales.

Aquí, como en la rúbrica precedente, la explosión y el desarrollo del Web ha modificado los componentes, al realizar un cambio a gran escala de las fuentes y de los recursos

⁸ Trad. Inglés: “Dirección y organización del conocimiento”

disponibles. De esta manera, la iniciativa lanzada por los arquitectos del Web tradicional para la creación de un Web semántico confluye felizmente con las aportaciones de los investigadores en este área.

Evolución

La definición del documento tradicional de acuerdo a esta dimensión podría formularse por medio de la ecuación siguiente: *Documento = inscripción + sentido*.

Aquí, el soporte es meramente accesorio, incluyendo también al documento tradicional, desde que la inscripción se preserve. Lo importante es el contenido, materializado por dicha inscripción, que funciona a la vez como vehículo de sentido. Éste, en sí mismo, se construye en relación al contexto de producción y de difusión del documento, que condicionará la interpretación del contenido.

Tres ideas de peso parecen conjugar esta dimensión, a partir del triángulo clásico en semántica. La primera se refiere a la creación de documentos; la segunda, a su interpretación y la tercera, a los signos que los constituyen.

“Pensar consiste en clasificar”; mediante la elaboración de documentos se aísla y se ordena el discurso para que nos ayuden a “pensar el mundo”. La organización del documento constituye una forma de construir o de traducir la forma en la que entendemos la sociedad. De esta manera, la noción de género textual y la de colección son fundamentales. En efecto, los documentos se reagrupan en grandes categorías cuyas unidades temáticas guardan entre sí cierta homología y relación. Esta operación se realiza a la vez en el marco de la organización del documento, que acabamos de señalar, así como en el de la estructuración de la colección. La clasificación varía según sean las diferentes situaciones y épocas, pudiendo estar sujeta a estrictos formalismos o ser simplemente implícita a ambas fases. Puede referirse a acciones muy precisas y organizadas (documentación de identidad, formularios, contratos, etc.) o, al contrario, a simples fórmulas de atención, impresiones, sensaciones (medios, ficciones, etc). La clasificación denota nuestra representación social y las diferentes lecturas que hacemos del mundo que nos rodea. Exige la existencia de un sistema que permitirá emplazar el documento en su conjunto y localizarlo al mismo tiempo. Se trata de un proceso de indexación en sentido estricto o figurado y, por lo tanto, de una serie de sistemas de clasificación concretos o abstractos.

Un segundo punto primordial es el que se refiere a la interpretación ¿Qué tipo de vínculos sugiere o instauro el documento y bajo qué forma lo hace? Un documento sólo tiene sentido por medio de la lectura y de la interpretación que un lector hace de él. Dicha interpretación depende en gran medida del contexto en el que se practica. Un mismo documento tomará sentidos diferentes, incluso opuestos, dependiendo de la época y del contexto individual del intérprete. Éste último, en cierta medida, recreará el documento cada vez que lo aisle y lo comprenda. El lector debe entenderse aquí en un sentido general; pudiendo tratarse tanto de una persona física, como de un grupo de personas en espacios y tiempos diferentes, o incluso de una máquina.

En la dimensión que aquí nos ocupa, un documento se toma en el marco de una doble relación, relación con el mundo documental (clasificación) y relación con el mundo natural (interpretación). Estas relaciones se realizan dentro de un “límite de espera”, que es el conjunto de signos que construyen el contrato de lectura entre el lector y el documento, permitiendo descifrar sin dificultades el sentido, que se situará ad hoc en su contexto de interpretación. Los editores, por su intervención en el contexto, su formalización y también por medio de una acción comercial, son los primeros artífices de esta construcción de documentos publicados. Por tanto, la noción de “contrato de lectura”, cuya importancia habíamos señalado en la rúbrica precedente relativa a la forma, recobra aquí una fuerza suplementaria, al ser indispensable para la comprensión del documento.

La tercera idea motriz se refiere a los propios signos. Todo objeto es potencialmente un signo y en consecuencia, podría convertirse en documento. Un tema de discusión, desde ahora clásico, muestra cómo hasta un antílope en un zoo (por tanto, en un sistema social de clasificación) tiene, en efecto, una dimensión documental. La gran parte de documentos se construyen a partir del lenguaje, fundamentalmente hablado o escrito. El zoo, en sí mismo,

se construye en función de un discurso; el antílope, de algún modo también, responde a un documento. Se podría hacer la misma observación en lo referente a aquellos documentos audiovisuales que se acompañan siempre de “leyendas de lectura”, representadas como una serie de documentos desde su fabricación hasta su explotación final. La estructura es la de la lengua escrita, desde la letra del alfabeto en las lenguas indoeuropeas hasta el discurso organizado; por tanto, la mayoría de los documentos. Compuestos de fragmentos, más o menos discretos, se pueden aislar con mayor o menor facilidad, así como distribuirse o analizarse, siendo objeto a la vez de reglas de sintaxis, estilo y retórica. Este uso de la lengua natural, viva, le confiere al documento una mayor plasticidad.

La explosión documental, es decir, la llegada irreversible de un considerable número de documentos desde finales del siglo XIX, ha dado lugar a la invención de los llamados “lenguajes documentales” (referencias bibliográficas, índices, tesauros, resúmenes, etc.). Organizados de forma asociada o jerárquica, se integran directamente en la tríada precedente: en efecto es posible construir un lenguaje artificial o formal a partir de los textos de documentos (de imágenes, o incluso de objetos) que permitan a la vez su clasificación dentro de un orden para facilitar a posteriori su búsqueda.

Hace ya tiempo que los archivistas realizan las colecciones a partir de los datos complementarios de los documentos y de sus productores, en el contexto de la descripción archivística, que considera la noción de contexto del documento como el elemento clave para su futura explotación en la red.

La construcción de estos lenguajes plantea numerosos problemas. En primer lugar supone una normalización, un cierto número de reglas comunes que sirven de consenso entre los diferentes protagonistas. No bastando dicho consenso, hay que añadir la motivación. Cada persona que participa en este esfuerzo, debe al menos recoger un beneficio claro, sin el que la construcción colectiva no sería eficaz. Por último, estos lenguajes se mueven entre lo contingente y lo universal. Tal ambivalencia no se comprende a menudo. No se trata de un vacío conceptual o de una incapacidad para elegir. Al contrario, estamos ante la dinámica fundadora del movimiento documental, que descansa sobre la tríada que sirvió de introducción a esta rúbrica: los signos forman parte de una dialéctica entre lo general (que clasifica) y lo particular, que refiere (interpreta).

La insistencia, justificada o no, de los documentalistas para diferenciarse de los bibliotecarios por el servicio específico que ofrecen: la búsqueda revela una concepción de la información específica, así como su desvinculación con el soporte. Los primeros se dedicarían a analizar el contenido de los documentos con el fin de presentar a los usuarios directamente las respuestas que precisan, y no simplemente el o los documentos que probablemente contendrían. Así, los documentalistas participan en la interpretación de los documentos al construir de alguna manera un informe documental adaptado a las necesidades del lector en cuestión.

Las “ciencias de la información” forman parte de este movimiento. El término de “información”, a medio camino entre “dato” y “conocimiento” no se ha definido como conviene. Tendríamos que referirnos, más bien, a unidades documentales. Las ciencias de la información intentan comprender cómo estas unidades se clasifican (una idea científica se presenta en un artículo, a la vez publicado en un número de una revista, que se difunde según un título o cabecera, formando parte de una colección, etc.). Igualmente, tratan de ver cómo se distribuyen, a partir de reglas estadísticas generales; cómo perfeccionar dichos lenguajes documentales y, sobre todo, cómo analizar por lo menudo el proceso de búsqueda de información entre un usuario o lector y un sistema de acceso.

En un primer momento, los documentalistas se sirvieron de lo digital únicamente como herramienta ventajosa, para la clasificación en unidades temáticas de los lenguajes documentales y en relación a una base de datos bibliográficos. Muy pronto, el tratamiento informático de la lengua coloquial, al igual que la producción y la gestión directa de documentos electrónicos; el éxito del Web y, para terminar, la modelización del razonamiento, cambiaron el componente práctico.

El tratamiento automático del lenguaje supera ampliamente la problemática documental. No obstante, al contrario, el tratamiento del documento comprende necesariamente los avances

así como las dificultades de las herramientas de tratamiento de la lengua, por cuanto éstas se vinculan al texto integral. Ya sea por medio de la indexación automática, ya sea por los resúmenes, o incluso por los sistemas informáticos de pregunta-respuesta, informáticos y lingüistas aúnan competencias sirviéndose de programas estadísticos y morfosintácticos. A su modo, han seguido el camino paralelo al de los documentalistas. Con la ayuda de filtros de cálculo reconstruyen, si no un lenguaje informático, por lo menos un texto que representa de manera estructurada el contenido del documento, permitiendo así un tratamiento automático por ordenador. Los resultados fueron menos probatorios de lo que esperaban sus conceptores en un primer momento, pues las mejores herramientas han sido aquéllas que integran una parte del trabajo humano, bajo la forma de útiles de ayuda, y no de tipo automático.

Sin embargo, para el internauta no iniciado, su eficacia como motor de búsqueda en el Web resulta asombrosa. Parece chocante la manera en que éstos albergan las tradicionales cuestiones biblioteconómicas, ya reformuladas a la vez en el campo de las ciencias de la información, quizás porque tales motores se dirigen a un gran número de documentos y de internautas simultáneamente: las leyes bibliométricas (Zipf), las colecciones (copias secretas), indexación y palabras clave (metadata), citas (vínculos), préstamos (hits) se renuevan ampliamente por la potencia del cálculo informático, utilizando para ello las aportaciones del tratamiento automático de la lengua, pero no siendo en la mayoría de los casos sino un bricolaje empírico de métodos, en lugar de un análisis científico riguroso.

Como en el enfoque precedente, los informáticos se han dedicado a aislar los elementos lógicos con la intención de modelizarlos. Pero aquí se han centrado exclusivamente en el contenido. Al igual que en el caso anterior, podríamos representar la transformación por medio de la ecuación: *Documento = sentido + inscripción* se convierte con lo digital en *Documento digital = texto informativo + conocimientos*. El cambio de *inscripción* por *texto informativo* significaría que tal texto (en un sentido más amplio, comprendiendo lo audiovisual) podría estar sujeto a un tratamiento que permitiría identificar las unidades de información. El cambio de *sentido* a *conocimientos* introduciría la noción de personalización del documento por un lector o usuario dado.

La llegada anunciada del Web semántico se entendería a la vez en esta continuidad de resultados y al menos como un avance metodológico, sino como una ruptura. Para la primera interpretación, se anotará el hecho de que la estructura XML, por ejemplo, es siempre formal; por lo que se refiere a RDF, ésta siempre se indexa. Se trata, desde esta óptica, de construir una biblioteca multimedia distribuida a la escala de la red y en la red, integrando para ello las herramientas de búsqueda más perfeccionadas. La ambición es pues mayor. Se plantea como objetivo pasar de un Web constituido simplemente por un conjunto de ficheros asociados a una red que utilice por completo las capacidades de cálculo de los terminales conectados, mediante los tratamientos automáticos del texto. Con esta finalidad, el uso de “metadata” que puede modelizarse y combinarse, resulta esencial. De este modo, los promotores del Web semántico crean una serie de lenguajes documentales que han denominado “ontologías”.

Era inevitable el cruce de los conceptores del Web con los investigadores en ingeniería de conocimientos, que persiguen la modelización del razonamiento. Los últimos reflexionan desde los años ochenta sobre el modo en que los razonamientos contenidos en documento, deben darse a conocer. Integran con especial interés las cuestiones relativas a los estatutos de los documentos, a la modelización de los razonamientos y, sobre todo, a las ontologías. Éstas se han definido como las representaciones posibles en una materia o área de conocimiento, acentuando la disociación (a menudo provisional) entre un razonamiento heurístico y la descripción de los conceptos que manipulan dichas heurísticas. La asunción de tal disociación constituía también un medio de simplificar la modelización de dos tipos de conocimientos, considerados independientes en un primer momento. Las ontologías focalizan los aspectos fundamentales de la materia (como la medicina o una rama de la medicina, por ejemplo), su vocabulario y luego el sentido que evocan. Este sentido se compone de dos facetas, la que entiende el ser humano, que es la semántica interpretativa, y la que comprende la máquina, que constituye la semántica formal de la ontología. Éstas

pueden verse como una estructuración más rica que la de los tesauros o los léxicos utilizados hasta ahora, al introducir por un lado una dimensión semántica (o red conceptual) y por el otro, en algunos casos, una dimensión léxica que mejora el acceso a los documentos. Pero una de las principales riquezas de las ontologías es, precisamente, su lado formal, que permitirá el uso por un programa informático allí donde el tesoro fracasaba. El aspecto formal se obtiene por medio de la descontextualización de los conceptos que se incluyen en la ontología. De ahí la necesidad, para facilitar la comprensión, de mantener y asociar esta ontología a la dimensión léxica a la que pertenece, los textos.

Así, al igual que para la entrada precedente, aunque sin duda de un modo menos avanzado, estamos quizás a punto de dar el paso a una nueva etapa para el documento digital, gracias a las aportaciones del Web semántico. Esta etapa podría representarse por la transformación de la ecuación presente: *Documento digital = texto informativo + conocimiento* que daría lugar a *WS = texto informativo + ontologías*.

Sin embargo, el aumento del número de documentos accesibles por una forma que no integra datos complementarios sobre el documento (“metadata”) es más importante que la de aquellos que se indexan de manera convencional. Aún peor, dada la gran concurrencia en el Web, aparecen nuevas estrategias oportunistas de indexación, con el propósito de confundir las búsquedas a partir de motores generalistas. Así, parece que, al menos en un primer momento, se encuentran dinámicas paralelas. Por un lado, para las comunidades autorreguladas, que se centran en el desarrollo de una investigación documental avanzada (dirigida a expertos, a empresas o a los medios de comunicación, etc.) Dichos documentos se indexarán desde su fabricación a partir de “lenguajes profesionales” valiéndose de técnicas manuales y asistidas al mismo tiempo. Por otro lado, los metalenguajes automáticos, más ligeros y ocasionalmente adaptados a los mecanismos de búsqueda por grandes categorías, continúan perfeccionándose por medio de sistemas cada vez más abiertos a los internautas.

Cuando se observa la evolución y los avances de las investigaciones en esta dimensión, se advierte un aspecto cíclico: los cambios de soporte, de jerarquía o de útil rencuentran las viejas cuestiones. La construcción, controvertida, de un lenguaje paralelo aparece a cada etapa. De esta manera, los partidarios de la etapa anterior tienen la impresión de que los recién llegados redescubren los problemas de antaño, mientras que éstos últimos sostienen que el paso realizado obliga a discutir de nuevo el conjunto de los problemas. No resulta del todo sorprendente el que la construcción de tal lenguaje sea así cíclico. A cada cambio de soporte o de jerarquía es necesario reconstruir su estructura. Deben ahora considerarse, además del conjunto de datos a representar, su aspecto multicultural y multilingüe. Al mismo tiempo, sus fundamentos no se discuten del todo, aunque deberían conocerse mejor y consolidarse más.

Definición 2

Según esta segunda dimensión, podríamos presentar una nueva definición del documento, sin pretender por ello todavía que englobe la noción en su totalidad. Esta definición debe tener en cuenta la capacidad de tratamiento del contenido ya sea para la investigación, ya sea para la simple localización del documento. Ésta se refiere al segundo capítulo del contrato de lectura que habíamos identificado, el de la inteligibilidad:

Un documento digital es un texto en el que los elementos resultan potencialmente analizables por un sistema de conocimiento para que un lector competente los explore.

Nuevamente, la definición es difícil. Recordemos las transformaciones de las ecuaciones más esquemáticas, pero también más nemotécnicas:

Documento = inscripción + sentido se convierte con el componente digital en *Documento digital = texto informativo + conocimientos* que podría dar lugar, con el Web semántico a la fórmula *Documento WS = texto informativo + ontologías*.

Incógnitas

Aproximándose a la comunicación humana, los investigadores de esta rúbrica han aumentado ampliamente el grado de complejidad de la problemática a tratar. Un buen número de cuestiones quedan sin resolver. Desde el punto de vista lingüístico, por ejemplo, cabría interrogarse sobre la aplicación de las herramientas utilizadas para aquellas lenguas cuya estructura y escritura no se corresponden con la indoeuropea. Por otro lado, la frontera entre la automatización y el trabajo humano intelectual sigue basculando.

Pero, para nuestra finalidad, anótese sobre todo el hecho de que para los investigadores que privilegian esta entrada, el documento resulta a menudo una noción secundaria; sólo el texto, el contenido, cobran especial interés. Sin embargo, el contenido, ya se ha dicho al comienzo de esta rúbrica, no tiene valor sino por su contexto ¿Acaso el documento no constituye precisamente una de las construcciones de este contexto, al situar las informaciones que contiene en relación a las que se contienen en otros documentos y permitiéndole al lector guardar así una indicación del valor del contenido por el estatuto del documento? Dicho de otro modo, al centrar exclusivamente su atención en el tratamiento del texto... ¿Acaso no se subestima el valor semántico de la disposición del documento? Una mejor consideración de la forma material se erige como centro de la problemática de los nuevos proyectos de investigación.. Lo más inmediato consiste en servirse de las balizas estructurales (y de su futuro semántico) para así modular los análisis de textos, la identificación de conocimientos o la anotación. Un estudio más minucioso se dedicaría a la integración de los elementos formales de tipo material, como es el caso de las pólizas, los quebrados, las escotaduras o las enumeraciones. En este sentido, la colaboración entre los especialistas del documento y la modelización de conocimientos resulta indispensable.

A partir de la tríada destacada como introducción de esta rúbrica, se plantean una serie de cuestiones:

¿Sería posible tratar el sentido de un documento sin establecer una relación, próxima o distante, con el conjunto al que se refiere (colección, categoría, renvío, bibliografía,...)? Dicho de otro modo ¿Cómo se podría integrar, más allá de un trabajo de laboratorio basado en corpus cerrados, el análisis del documento como “eje de la red”, generador de estructuras que trabajan el sentido? Es la misma incógnita, aunque renovada por lo digital, basada sobre el enriquecimiento a partir de vínculos entre los documentos, en las nuevas situaciones de hipertextualización o de la creación intencionada de una colección.

¿Puede validarse una información por otro elemento distinto a la autenticación del documento que la contiene?

El problema de la confianza constituye en la actualidad un tema estudiado por los investigadores en informática (representantes de conocimientos, entre otros) que se interesan en el Web semántico y por el consorcio W3C. Buscan una solución técnica (añadir una capa formal que pueda interpretarse por los programas agentes) que permita al mismo tiempo a los usuarios / lectores de un “site” el reforzar o disminuir la credibilidad o la confianza que se les puede acordar a las informaciones que contenga. Esta visión técnica del problema enlaza con la idea de que son los internautas los que validan y popularizan la información de un “site” Web. El proyecto sería pertinente en el seno de una comunidad especializada y que posee sus convenciones, viéndose limitado en el marco del Web. El problema es mucho más complejo que el del simple “voto” que plebiscita un “site” y lo acredita. Los enfoques sociológicos se imponen en este caso.

Cuando se trata de analizar el contenido de un documento con el fin de crear modelos de conocimiento para usos particulares, la validez y la pertinencia del documento se critican por otras fuentes de conocimiento, como son los expertos en un área o los usuarios. En función de los objetivos en cada caso, los propios métodos de extracción de conocimiento a partir de los textos determinan un peso igual o incluso mayor al de la formulación oral de tales conocimientos. El valor añadido por la autenticidad, la certificación y el reconocimiento del texto puede dejarse al margen en algunos casos.

¿En qué medida se puede aislar un elemento significativo en un conjunto que guarda una unidad de sentido, el documento como un todo? ¿Esta unidad no tiene a menudo un peso decisivo en la significación de los elementos que la componen? ¿Cómo se puede tomar en consideración la significación global, la unidad semántica de un documento, aprendiendo sin embargo sólo sus partes fragmentadas? Las incógnitas aquí planteadas triplican ampliamente aquellas que surgen en torno a los textos en semántica. El paso del texto al documento merecería sin duda un análisis más detallado.

Estas cuestiones tienen por supuesto respuestas diferentes según el tipo de documentos a los que se apliquen. Pero falta por el momento un avance real sobre la tipología y siguen siendo, a nuestro parecer, demasiado abiertas.

Documento como medio

Renovemos de nuevo con precaución la precisión del vocabulario escogido: el término “medio” debe entenderse aquí en un sentido amplio. Reagrupa todas las aportaciones que analizan el documento como un fenómeno social, un elemento tangible de una comunicación entre personas humanas. Esta rúbrica etraña un análisis de la documentación, una comunicación particular en la que el documento se entiende como el vector o vehículo de un mensaje entre las personas. Podría afirmarse así que se trata de una tercera dimensión del contrato de lectura, la de la sociabilidad.

Especialidades relacionadas

Anótese primero que el ámbito social que concierne podría subdividirse en dos: por un lado, figuran las organizaciones que utilizan los documentos para su regulación interna, y con el fin de alcanzar los objetivos que se han fijado en un principio; por el otro, están las sociedades o las colectividades abiertas en las que circulan los documentos.

Sin duda, todos los investigadores anteriores podrían incluirse en esta categoría, ya que todos se interesan en una actividad social; sin embargo, aquí clasificaremos aquéllos cuya entrada es en primer lugar social, y no instrumental.

De igual modo, en lo referente a las prácticas tradicionales, los oficios anteriormente citados se encuentran de nuevo aquí, pero nosotros insistiremos en el caso de los archivistas, cuya primera misión consiste en la conservación de la huella de la actividad humana, al guardar los documentos producidos a lo largo de los años, así como en el de los editores, cuya tarea se centra en favorizar la construcción de dichos documentos, así como en publicar aquéllos que pudieran interesar a una colectividad.

Las disciplinas de las ciencias humanas y sociales que se interesan en los intercambios se encuentran potencialmente relacionadas con esta dimensión. De hecho, los sociólogos, los economistas, los juristas, los historiadores, algunos psicólogos, un gran número de filósofos y, por supuesto, los investigadores en ciencias de la comunicación, en ciencias políticas y en ciencias de la gestión se interesan, directa o indirectamente, en los documentos a partir de su rúbrica disciplinaria.

Lo digital ha renovado el interés de numerosos investigadores de estas disciplinas, tanto en el ámbito global como en las situaciones particulares. Así, aun no asumiéndose en todos los casos, existe una relación entre la reflexión sobre el documento y el nuevo interés hacia las comunidades relacionadas, de tipo colaborativo, asociativo, de trabajo en red, la memoria y el patrimonio, la propiedad intelectual, etc.

Pero en esta entrada el desfase entre informáticos y el resto de investigadores es todavía mayor. Muy pocos especialistas en ciencias sociales y humanas tienen un conocimiento real de la informática. En sentido inverso, los informáticos tienen a menudo una comprensión vaga de la problemática social. Esta zanja conduce por momentos a entusiasmos fascinantes o, al contrario, a rechazos radicales, tanto para los partidarios de las ciencias sociales y humanas, como para los ingenieros.

Evolución

Un documento da un estatuto a una información, a un signo que se materializa. Un grupo lo suscita y lo porta, lo conserva y lo materializa. Como hemos sugerido en la introducción de este texto, constituye una prueba que da fe de un estado y es un anuncio que prevé un acontecimiento. Se trata, pues, de un discurso cuya firma lo vincula a un autor. Se trata de un testimonio, aun no habiéndolo pretendido en el momento de su concepción. Es una de las piezas constituyentes de un informe.

Para armonizar con las rúbricas anteriores, proponemos una tercera y última definición que consiste en la formulación de la ecuación: *Documento = inscripción + legitimidad*. Nos parece que esta ecuación permite representar el proceso social de puesta a punto del documento. El estatuto de documento se adquiriría a través de dos condiciones: la inscripción debe superar la comunicación íntima (entre algunas personas privadas) para llegar a legitimarse, y ésta, por su parte, debería superar también lo efímero (trascender el momento de su enunciación) y en consecuencia grabarse o inscribirse. Tales condicionantes implican que si todo signo es potencialmente un documento, un signo particular no lo es necesariamente (incluso de responder a las dos dimensiones señaladas anteriormente). Por ejemplo, un periódico no constituye un documento, si alguien no toma la iniciativa de hacerlo público o al menos de difundirlo más allá del círculo restringido de familiares del autor. Incluso, una emisión de radio o de televisión en directo sólo alcanza el estatuto de documento a condición de que alguien la haya grabado para una utilización social en el futuro.

Esta posición no constituye la unanimidad completa entre los creadores de este texto. Para algunos, el valor de un documento podría preceder su intercambio o su grabación.

El estatuto de documento no se adquiere para siempre; puede darse y también perderse en el olvido colectivo y definitivo e incluso retomarse si alguien descubre y legitima de nuevo un documento que ha desaparecido de la conciencia colectiva, pero que no se ha destruido.

Sin embargo, la ecuación no considera la función social de los documentos. Éstos se utilizan en la regulación de las sociedades humanas, asegurando una comunicación y una permanencia de la norma y de los conocimientos necesarios para garantizar su pervivencia o su continuidad. En cierta manera, podría decirse que el contrato de lectura en el que habíamos identificado dos dimensiones correspondientes a las entradas precedentes: la legibilidad y la comprensión, adquiere aquí su tercera dimensión: la sociabilidad, una apropiación mediante la que el lector adquiere el conocimiento sobre el documento y resalta así su participación en una sociedad humana o, a la inversa, la inscripción sobre un artefacto de una representación del mundo natural y su inserción en el patrimonio colectivo.

Un documento no se publica necesariamente. Gran número sólo se consultan por un número de personas, por ejemplo en el caso de temas privados (historial médico, transacciones privadas) o porque contienen secretos no divulgables. Sin embargo, guardan un carácter social en el sentido en que están redactados de acuerdo a las reglas establecidas que fundamentan su legitimidad, porque son utilizados en las relaciones formales y porque, en caso de disfunción, seguirán teniendo un valor de referencia. A la inversa, la publicación, amplia o limitada, constituye el sistema más simple de legitimación. En efecto, una vez publicado, es decir, consultado en potencia por un número considerable de personas, el texto forma parte del patrimonio común. No puede modificarse sin topar con dificultades en el proceso; su valor se aprecia de forma colectiva.

La multiplicación de documentos está vinculada a la evolución de las sociedades, en función de dos dinámicas, externa e interna, que se reconfortan recíprocamente: la del uso social de los documentos, en primer lugar y en segundo lugar, la de la economía propiamente.

La organización política y social se sustenta en la producción y en el intercambio de documentos. Las religiones y sus representantes clericales, los estados y sus respectivas administraciones, las organizaciones productivas y el comercio, la sociedad civil en sus diferentes manifestaciones; su evolución histórica, su geografía y sus propias culturas; sus funciones, en cambio permanente, se sirven, todavía hoy, y proporcionan a los documentos sistemas para su regulación interna así como para la afirmación concurrente de su identidad

y posición. Así, citemos las principales fuentes de la actividad documental en los países occidentales sin ser exhaustivos:

- En Francia, el paso del Antiguo Régimen a la República, después, el del Estado militar al Estado Providencial y, por fin hoy, del Estado a superestructuras más amplias, como Europa o la mundialización no se ha llevado a cabo sin consecuencias en la producción de los documentos, su función y su número. Basta, como comparación, el evocar la importancia del documento en la historia paralela de la administración en China para apreciar cómo éste resulta a la vez fundamental y, sin embargo, muy específico en cada civilización.
- La industrialización con todos sus saberes y las normalizaciones técnica, organizacional, transaccional y contable que la han acompañado, produjo un número considerable de documentos. Quizás sea éste el factor determinante en el fenómeno de explosión documental que hemos citado anteriormente.
- Los progresos científicos y los relativos a la educación han aumentado considerablemente el número de productores y de consumidores de documentos para el funcionamiento interno de la ciencia y, más aún, para la popularización del sinfín de saberes parciales que la acompañan.
- Los intercambios comerciales o no comerciales que han estallado a raíz del desarrollo de los transportes, de las telecomunicaciones y de la apertura de las fronteras se sirven de numerosos documentos para “fluidificarse” (materialización de las transacciones, notas técnicas que acompañan los productos y los servicios, informaciones comerciales, etc.).
- El desarrollo del tiempo libre, el aumento de la esperanza de vida, el crecimiento del “espacio público” son todavía factores del desarrollo de la cultura, así como de uno de sus principales vectores: el documento.

Esta dinámica ha sido el objeto de descubrimientos particulares, pero las pocas tentativas de aprensión general nos parecen ensayos más próximos a la especulación que a la demostración, quizás porque constituyen por el momento trabajos aislados.

La segunda dinámica que permite considerar el documento como medio es la de su economía interna, que se construye a partir de la evolución tecnológica que la constituye (evolución desarrollada en las dos rúbricas anteriores) y, por otra parte, por medio de las modalidades de la puesta a punto del documento. Éstas suponen, en efecto, un trabajo en el que hay que encontrar los medios para su realización. La disposición del documento puede analizarse como un acto de comunicación ordinario con uno (o varios) remitentes, por un lado, y con uno (o varios) destinatarios, por el otro. Los oficios se han especializado en uno u otro momento del proceso, o en un campo de aplicación específico. Los sistemas se han construido y formalizado para responder a la regularidad de la producción. Los empresarios, grandes o pequeños, se han lanzado a la aventura y las organizaciones se han hecho cargo también de esta nueva situación. Estos dispositivos tienen un coste de puesta en marcha, de mantenimiento y de inercia.

Dos corrientes de investigación principales se consagran a estudiar la economía de esta disposición del documento. La primera se interesa en la comunicación organizacional y estudia sobre todo los documentos en el proceso de trabajo; la segunda, analiza la comunicación de los medios y se centra en el proceso de publicación.

La comunicación en el seno de la organización estudia el documento en el seno de las prácticas profesionales caracterizadas, y, por tanto, limitadas por sistemas de reglas. Lo conserva a varios niveles: primero, como un escrito identificado a un contexto, formalizado por reglas de escritura, de circulación, de uso, que inscribe una intencionalidad ligada a una acción, y guardando la huella de las negociaciones sociotécnicas que se llevan a cabo en torno a él. Esto conduce especialmente a estudiar los procesos de fabricación y de gestión de los documentos, actividades que no son solamente propias de los actores especializados, sino que se distribuyen en el conjunto de la organización. Seguidamente lo analiza como un elemento estructurante de la organización en calidad de soporte de coordinación. Por fin, lo advierte como un medio utilizado por los actores individuales o colectivos en el marco de sus diferentes estrategias. En el ámbito metodológico, el documento se valora como

elemento observable, que permite estudiar al mismo tiempo las relaciones entre los actores (se trata de un mediador), los modos de regulación (es una herramienta de dirección) y las recomposiciones en el seno de la organización (al revelarlas).

Los primeros avances del análisis sólo tocan una parte de la actividad documental y están todavía lejos de formalizar los límites de una economía general de los documentos. Quedan aún muchas incertidumbres por desvelar; por ejemplo, en lo tocante a las relaciones entre los sistemas documentales y los sistemas de la organización; la evolución de las diferentes profesiones, la mediación o incluso la economía de las bibliotecas o de los archivos. En el ámbito de los archivistas, las discusiones se centran en las prácticas de *Records Managements* y del *Business Reengineering*, que están normalizándose. La doctrina actual (lejos de asimilarse por los productores documentales institucionales) quiere que las misiones (objetivos institucionales) engendren procesos (funciones de tipo organizativo) que den lugar a procedimientos (métodos de acción formalizados) que se materialicen en documentos (o transacciones).

En relación a los medios, la economía de varios sectores se conoce porque constituye el objeto de análisis particulares. Cítese la comunicación científica que tiene el papel de encargarse de la publicación de los artículos, de la revisión realizada por los colegas, las citas o los borradores previos a la publicación. Hasta los medios de masa, dirigidos a un público más generalista, que establecen una jerarquía entre una edición artesanal, sustentada por la venta individual de objetos, la dialéctica entre los fondos y el *best-seller*; las redes de distribución; la radiodifusión, más industrial, y organizada en función de la captación a domicilio de la atención del destinatario, que a la vez se vende a los anunciantes interesados.

Se ha profundizado en otros temas, teniendo en cuenta las apuestas económicas que representan como el derecho de la propiedad intelectual. En este campo, las tradiciones diferentes entre el derecho de autor latino y el *copyright* anglosajón permiten distinguir algunas de las propiedades del documento. El derecho de autor privilegia la vinculación del autor a su obra, mientras que en el segundo caso, se antepone la noción de publicación, otorgando la propiedad intelectual al que toma la iniciativa. En cierta medida, podría afirmarse que el derecho de autor es un derecho de la obra, mientras que el *copyright* es un derecho del documento.

Anótese un último punto en relación a la economía de los documentos, cruciales en nuestro propósito: cuanto más conocido es un documento, más se leerá y cuanto más se lea, más se conocerá. A partir de las relaciones entre los lectores y los documentos, se puede desarrollar un fenómeno de resonancia, que cobra diferentes formas y diferentes nombres según los sectores o las especialidades. Los profesionales del marketing o las estrategias de los medios de comunicación lo utilizan regularmente construyendo actualidades que revenden en otros soportes. En la comunicación científica, el factor de impacto, basado en el recuento de citas en los artículos, entraña el mismo proceso (y conduce a los mismos puntos). Esta cualidad es quizás la explicación de un gran número de características de la distribución de los documentos: *best-sellers* en la edición, *prime-time* en la radio y en la televisión, modos diversos, concentración y explosión o incluso regularidad casi perfecta de las leyes bibliométricas cuando grandes cantidades de documentos resultan accesibles de manera igual por un gran número de solicitantes.

Lo digital se traduce por movimientos contradictorios cuya interpretación no es fácil. La primera constatación sería la de la supresión, difícilmente medible, al efectuarse en orden disperso, de un buen número de documentos que, guardados en una forma tradicional, dan cuenta de los procedimientos. La difusión de las herramientas informáticas se tradujo a menudo por la disociación de funciones asumidas hasta entonces por un único tipo de documento. Tal es el caso de los registros del estado civil, que continúan guardándose en papel por razones jurídicas, aun cuando se digitalizan para facilitar su consulta. Cada vez más a menudo, y coincidiendo con la desaparición de puestos intermediarios, la sustitución es total: formularios, fichas y tablas, manuales, instrucciones que constituían los mejores momentos de la burocracia pública o privada se han remplazado por bases de datos y de intercambio de datos informáticos. Este movimiento que se bautizó hace algunos años como

“informatización de la sociedad” amenaza con acelerarse todavía más con los desarrollos destacados en las dimensiones precedentes.

Pero, en el mundo de las organizaciones, se constata de modo conjunto un aumento de una escritura y una disposición documental amplificadas de forma exponencial por la gestión de la calidad. El documento es el portador también de normas sociales y de tipo organizativo, lo que lo sitúa a la vez tanto como un soporte para la acción, como una memoria de las relaciones.

El hecho de integrar los datos o procedimientos en bases de datos no suprime su valor prescriptivo, al contrario. Los sistemas de Intranet, por ejemplo, confieren a los documentos un estatuto (de referencia, de útil, entre otros) asociándoles reglas de identificación y de circulación modelando y anticipando al mismo tiempo los usos posibles. Amplifican la visibilidad de las decisiones y de las actividades, haciéndolas ampliamente accesibles.

En esta perspectiva, menos que nunca, el documento no aportaría sentido alguno de manera aislada, sino que se constituiría por el registro informático de transacciones anteriormente definidas. La fijación de informaciones, efímera y necesariamente dependiente de tecnologías evolutivas, no constituiría en sí y solamente el documento; debería validarse por procedimientos certificados. Numerosos documentos pueden apropiarse de la transcripción de los procedimientos o de alguna de sus etapas: así, tales documentos no podrían finalmente entenderse sin la función del modo de traducción informático cuyos procedimientos de transcripción constituyen el objeto. Si a ello se añade los progresos en la firma electrónica, gran cantidad de transacciones podrían realizarse en el futuro sin los formalismos adoptados por el soporte papel.

Esta mutación acrecienta de modo considerable las posibilidades de control debido las facilidades de aumento de las informaciones. En el ámbito social, Francia se ha dotado de una protección legal, aunque a la vez frágil, dado el desarrollo de las operaciones digitales, por la ley “Informática y libertad”. En el campo económico, algunos analistas han visto en ello la emergencia de una nueva economía en la que el estudio de las contradicciones y desengaños supera largamente el marco de este artículo. Retengamos para nuestro propósito la idea de cambio radical de las estructuras socioeconómicas. Al igual que la era industrial, marcada por el intercambio, la sociedad de la información se caracterizaría por la posibilidad de reutilizar la información.

Así, desde el punto de vista de la comunicación en la organización, habríamos identificado una primera evolución en nuestra primera ecuación $Documento = inscripción + legitimidad$ hacia $Documento digital = texto + procedimiento$. Pero esta última ecuación no da cuenta de otro importante movimiento en curso referente a los medios y a partir de la llegada del Web.

El Web ha permitido proyectar lo digital en el conjunto de la sociedad de forma brutal. Para comprender el éxito que ha encontrado, por medio de una difusión explosiva en las poblaciones en función de los tipos de actividad, hay que retomar el espíritu que sustenta su arquitectura. La organización del Web, conforme a las orientaciones de los conceptores de Internet, se ha imaginado como una red de comunicación de varios a varios en el que cada rama, grande o pequeña, debía disponer de las mismas herramientas y erigirse a la vez como productor y consumidor. El Web presupone una concepción social, o mejor aún, de la comunicación social, próxima a la “República de las ciencias” o del movimiento de los programas abiertos. En el seno de semejante sociedad, cada persona es actor y responsable ante la comunidad de sus actos. De trasladar esta noción a nuestro ámbito de estudio, diríamos que cada uno es capaz de leer o de escribir documentos que conciernen la vida colectiva y que cada uno se encargará de publicar sólo aquéllos que beneficien y enriquezcan a la colectividad. Los genios pioneros del Web, al articular Internet y el hipermedio, construyeron un sistema a su imagen o más bien a imagen de la comunidad informativa a la que pertenecían. Esta idea está presente en un gran número de discursos y de iniciativas en el campo, empezando por el consorcio W3C. La industria del contenido (industria de programas y de las telecomunicaciones) no sin debates, batallas y compromisos, permanece atenta a los desarrollos que reconfortan sus posiciones, ya que

favorecen el aumento del tráfico y de los tratamientos en detrimento de una industria del contenido.

Pero todo el mundo no puede comunicar a todo el mundo; estaríamos pues ante una cacofonía, hay que hacerlo mediante representantes. Hasta ahora esta dificultad se resolvía por uno o varios sistemas de filtros que permiten seleccionar los autores pertinentes y configurarlos documentos que son representativos y útiles. Estos sistemas tienen un coste que no sabría diluirse en el funcionamiento comunitario, al haber desaparecido la igualdad entre los actores. Algunos de ellos sólo escriben en nombre de otros y los mediadores profesionales organizan el conjunto del sistema de publicación y de acceso. El sistema editorial es un avatar de esta organización, compromiso entre intereses privados y públicos. Se trata, en consecuencia, de un malentendido, voluntario o inocente, en un principio entre los sistemas concebidos por los pioneros de Internet, y reconfortado por los industriales del contenido y la realidad ordinaria de la comunicación social. Este malentendido es, sin embargo, fértil, en el sentido en el que permite a las colectividades en las que la comunicación se aferra al sistema tradicional encontrar un espacio para el intercambio. Proporciona también a numerosas instituciones, comenzando por aquéllas representativas del interés general, una herramienta simple para comunicar con la población. Así, el Web constituye un amplio bazar en el que se pueden encontrar multitud de documentos a consultar gratuitamente por el lector, y que están asociados entre ellos.

Algunos consideran que no se trata de una organización provisional, que ilustra en efecto la corta edad del medio. Este análisis no calibra en justa medida la ruptura que produce el Web. Es probable también que el filtro o la selección ya no se haga en el Web *a priori* como en los medios tradicionales, pero *a posteriori* y conforme a un sistema de “percolación” en el que los documentos más pertinentes se identificarían y se valorarían progresivamente en función del número de vínculos y mediante el juego de los motores. La red en sí sería lo importante, al permitir, por un movimiento continuo (los vínculos que nacen y desaparecen, los motores que buscan, las páginas que aparecen y desaparecen) la localización de los documentos. La implicación de un número consecutivo de internautas, hasta aquí distanciados del ámbito cerrado de los medios, y el éxito manifiesto del media en las prácticas aportan consistencia a esta hipótesis al procurar una dimensión y una rapidez inédita a la dinámica ordinaria de la legitimidad por medio de la notoriedad.

Siguiendo un razonamiento paralelo, pero en una perspectiva literaria, varios investigadores y ensayistas han visto con la llegada del Web y específicamente en las técnicas de hipertextualidad o de hipermedia, una supresión de los documentos. Así, la tríada clásica autor-obra-lector, que constituye el origen de los documentos literarios, dejaría lugar a un proceso interactivo en el que los vínculos entre las páginas accesibles jugarían un papel más importante que el texto, tal y como se construía por el autor hasta el presente. No obstante, a pesar de interesantes experiencias en la escritura hipertextual, que han revelado consecuencias semánticas y cognitivas significativas, todo parece indicar que el desarrollo del Web ha derivado en la multiplicación exponencial de los documentos electrónicos. Los vínculos entre las páginas parecen estructurarse de manera progresiva para construir nuevas normas de paratexto, reforzando, al contrario, el aspecto documental del Web.

Podríamos resumir nuestro desarrollo en el Web mediante la transformación de la ecuación inicial que daría lugar a *documento Web = publicación + acceso localizado*. La publicación en sí no determina la legitimidad; habría que añadirle la notoriedad, a partir de una localización del acceso.

En coherencia con nuestro razonamiento, los medios tradicionales no han llegado a construir en el Web modelos económicos viables. Sólo algunos sectores, que ya mostraban afinidades con las redes han encontrado modalidades de financiación. Es además posible que la música, a raíz del intercambio entre internautas esté redefiniendo su modo de distribución y de valorización. A la inversa, lo digital ha permitido reforzar los “antiguos” medios, de edición y audiovisual, permitiendo obtener ganancias de productividad al favorizar las sinergias y las diversificaciones. Así, a partir de una misma base de datos, por ejemplo, un periódico puede dar cuenta de la actualidad a la vez en soporte papel, en el Web, en la radio, por SMS, en audiotel, etc. Incluso, cada medio puede hacer valer sus

propios ámbitos de excelencia (notoriedad para la televisión o la radio, interactividad para el Web y el teléfono, apropiación para la edición) y por medio de la resonancia mencionada más arriba, conducir a beneficios inéditos. Estos cambios recientes quedan aún por evaluar. Ellos conducen también a grandes inversiones en las que los retornos se estipulan a medio plazo aun cuando el futuro es incierto. Tras una primera fase de encantamiento y de marcha a tientas, el Web no aparece sino como un medio suplementario del que conviene comprender las cualidades propias con el fin de articularlo con los medios existentes. Los desarrollos anunciados del Web semántico dejan presagiar otros desarrollos especialmente en las relaciones entre documento y servicio. Pero esto constituye aún una prospectiva difícil a analizar en esta rúbrica de fuerte dimensión social.

Definición 3

En esta perspectiva, habíamos pues identificado movimientos fuertes, en ocasiones divergentes y a menudo caóticos. No es simple en este estado el proponer una dimensión que refleje claramente esta tercera entrada. Por ello, nos aferraremos a una definición muy general:

Un documento digital constituye la huella de las relaciones sociales reconstruida por los dispositivos informáticos.

Al retomar las ecuaciones que habíamos reconstruido y transformado: *documento = inscripción + legitimidad* se convierte en *digital = texto + procedimiento* y *documento Web = publicación + acceso localizado*.

A pesar de las dificultades para la construcción de una definición, subrayaremos la importancia del tercer capítulo y del contrato de lectura identificado, el de la sociabilidad.

Incógnitas

La primera serie de incógnitas concierne la noción de archivo cuya base es la del registro y la conservación de documentos. El papel del archivo es el de preservar la memoria de una actividad humana. Una nueva tarea, más activa, emerge en este sector gracias a la digitalización. Archivos abiertos, recuperación de programas audiovisuales a partir del código fuente o de la difusión de emisiones de televisión; funciones de archivo en el Web y multitud de actividades inéditas se desenvuelven, mientras que las prácticas de la archivística se renuevan, la llegada de *records management* en las organizaciones aparece como condición para proceder a un archivo electrónico. Se sienten todavía un buen número de preguntas sin respuesta definitiva, en relación a una nueva función que se debe asumir: dudas entre el testimonio de una acción pasada y registro de una acción en curso; confusión entre el archivaje y la publicación; simple registro o preparación para una utilización futura. Más aún, ¿cómo preservar la memoria de un movimiento continuo de renovación de páginas asociadas entre ellas?

La segunda serie de cuestiones se refiere a la noción de atención (tipos de intenciones y de percepciones), sin la que un documento no encontraría un lector. La atención humana es limitada, por el tiempo disponible, por la fatiga del lector o por las competencias técnicas e intelectuales que dispone. Los radiodifusores conocen bien esta problemática.

El internauta, aun siendo necesariamente activo, no se “capturaría” como el auditor de la radio o el telespectador. Dicho de otro modo, el Web aúna la libertad de elección de la edición con la accesibilidad de la radiodifusión, o extiende los servicios de la biblioteca a todo el planeta, para la colección, y a domicilio, para la consulta. Así, las leyes bibliométricas y los efectos de resonancia amenazan con jugar a una escala inédita por sectores: la atención se concentra en un número reducido de documentos y se dispersa en un número más extenso. Estos fenómenos y sus consecuencias son por el momento poco estudiados.

Por otro lado, la intención de los promotores del Web consiste en ofrecer por igual y para el conjunto del planeta los “sites” y los documentos. Pero la difusión de las innovaciones son

muy desiguales. El Web y lo digital no escapan a la regla general. Peor aún, todo indica que se trata del medio menos compartido entre las poblaciones en el interior de cada país.

El tercer conjunto de incógnitas alude al olvido en la financiación del contenido. La ley del mínimo esfuerzo aplicada a la accesibilidad del Web hace que el internauta prefiera evitar todos los obstáculos y barreras durante la navegación, en lugar de afrontarlos. Así, escamoteará toda demanda de financiación directa. En la misma dinámica, un movimiento militante convierte lo gratuito en una especie de calidad natural del Web o en una accesibilidad al saber y a la cultura liberada de las limitaciones comerciales.

Oportunismo y política se cojugan para que, de modo progresivo, la economía del contenido en el Web se configure como un mercado institucional, del “B2B”⁸ ¿Es de verdad seguro que esta estructura de financiación garantice a medio plazo la diversidad de la pluralidad de los documentos electrónicos? ¿Es simplemente cierto que contenga los recursos suficientes, sector por sector, para alimentar la producción y la gestión de documentos?

El objetivo de traducir el conjunto de los existentes de un soporte tradicional a un estado digital está fuera de alcance. El desarrollo explosivo de la informática obliga sin embargo a hacer frente a tratamientos de gran amplitud, a excepción de acomodarse en una amnesia radical de nuestra cultura documental. En un futuro próximo será necesario estar en condiciones de hacer elecciones razonadas (¿qué hay que digitalizar en prioridad?) así como de construir las herramientas capaces de tratar a un coste razonable un gran número de documentos.

Conclusión

Las tres entradas que se han abordado han permitido sacar a la luz varias temáticas fundadoras del documento, reconfortadas o contestadas por su versión digital. Queda por llevar a cabo una síntesis que proponga una clarificación general, que englobe los tres puntos de vista, al igual que la paleta del pintor, que se constituye a partir de tres colores básicos. De forma más académica, ¿es posible proponer una teoría del documento que permita ponderar mejor las consecuencias actuales y futuras de lo digital?

Destáquese en primer lugar, sin sorpresa, cómo hemos identificado en cada una de las entradas diferentes etapas en la historia de la digitalización de los documentos que podemos ahora hacer madurar en paralelo. El documento tradicional descansa en un soporte, un texto y una legitimidad. Una primera fase de digitalización, en la que nos encontramos todavía, ha permitido mostrar estructuras internas, destacar la importancia de los “metadatos” para el tratamiento, así como la dificultad para su validación. Una segunda fase, ya iniciada, pero cuyo desenlace permanece incierto, se centra en el formato XML, que integra la estructura, pero no la forma, y se sustenta en las ontologías para encontrar y reconstruir los textos, promoviendo el acceso personalizado. Hay en esta evolución general un sentido del que convendría entender la orientación, las consecuencias y los límites.

Destaquemos a continuación que la oposición entre el papel y lo digital resulte vana. Casi todos los documentos actuales han existido previamente en formato digital y los que no confirman esta regla se arriesgan a caer en el olvido. A la inversa, un gran número de documentos digitales acaban imprimiéndose en un momento u otro, ya sea por una impresora individual, ya sea por una de tipo profesional. De este modo, lo importante consiste en determinar la noción de documento general, del que lo digital es a la vez un elemento revelador y un factor de evolución.

Anotemos por fin que en cada rúbrica hemos insistido en la idea del contrato de lectura traducido por la legibilidad en la primera, por la comprensión en la segunda y por la sociabilidad en la tercera. Es probable que este contrato a tres facetas presente, en cada uno de los matices expuestos, la realidad de la noción de documento. Un documento sólo sería finalmente un contrato entre los hombres cuyas cualidades antropológicas (legibilidad-comprensión), intelectuales (comprensión-asimilación) y sociales (sociabilidad-integración)

⁸ *Business to Business*. Trad. Inglés: transacciones entre empresarios

fudarían una parte de su humanidad y de su capacidad para convivir. En esta perspectiva, lo digital constituye sólo una modalidad de multiplicación y de evolución de estos contratos. Pero la importancia que ha tomado, su perfección y la rapidez de su difusión exigen ahora más que nunca el llevar a cabo un análisis preciso y justo. Hemos mostrado claramente, en particular en lo referente a las incógnitas presentadas, que ninguna de las entradas era independiente. Sería absurdo intentar separarlas; al contrario, la noción no se esclarece sino por su superposición. Pero hemos constatado también que cada una se tomaba a partir de un movimiento de investigaciones pluridisciplinarias que guardaba su autonomía y cuya especialización implica experiencias muy complejas para compartirse realmente. Este texto es una llamada para profundizar tanto en cada enfoque como en su cruce.

Rogelio T. Pedauque
CNRS-STIC
08-07-2003