
Metodología para el procesamiento de la información a partir de fuentes indirectas: el caso del Quijote

Methodology for information processing from indirect sources: the example of Cervantes' Quixote

Aitana MARTOS GARCÍA (1) y Agustín VIVAS MORENO (2)

Grupo de Investigación ARDOPA (Archivística y Documentación del Patrimonio), Universidad de Extremadura. Facultad de Biblioteconomía y Documentación (1) amargar@alcazaba.unex.es. (2) avivas@alcazaba.unex.es.

Resumen

Se pretende establecer los criterios metodológicos necesarios para extraer el valor informativo a partir de fuentes de información indirectas —tales como memorias y relatos de viajeros u obras literarias con el objeto de confeccionar un corpus documental e informativo sobre la Cultura Escrita y la historia de la lectura. Para ello, se combinan técnicas de análisis de contenido, documentales y de aplicación de las Nuevas Tecnologías de la Información. El trabajo se estructura en diversas fases: procesamiento previo de las fuentes originales, a través de su análisis intuitivo y conceptual; extracción de los fragmentos textuales o referencias; construcción de las herramientas documentales específicas para ellos (incluyendo un microtesauro de descriptores) y la posterior indización de las referencias; y, por último, integración de todo lo obtenido dentro de un gestor documental automatizado que localice dichas referencias de acuerdo a las necesidades del usuario. Como ejemplo, se mostrará un gestor documental elaborado para extraer toda la información relativa a la Cultura Escrita dentro del Quijote. Este estudio nos permitirá concluir que, a partir de los criterios metodológicos expuestos, es posible extraer el valor informativo tanto explícito como implícito en estas fuentes indirectas y servirla al usuario de la forma más rápida y eficaz posible.

Palabras clave: Cultura escrita. Fuentes indirectas. Gestor documental. Procesamiento de la información. Quijote. Tesoros.

1. Introducción

La investigación científica, y aún más la histórica y la cultural, parte del estudio de muchas fuentes que a menudo están dispersas. A veces se condensan pequeñas cantidades de información en un volumen amplio o viceversa; y en otras ocasiones, la información aparece en fuentes originales que pasan desapercibidas para el investigador. La recopilación y extracción de la información de esas variadas y desperdigadas fuentes consume un gran volumen

Abstract

This paper intends to set up some methodological guidelines required for extracting the information value from indirect information sources - as memories and travelers' tales or literary works. The aim is creating a documental and informative corpus about written culture and history of reading. To reach this aim, several methodologies are used, combining analysis of texts, information science procedures and application of Information Technologies. The project had the following phases: previous processing of the original sources by means of their intuitive and conceptual analysis; extracting of the textual passages or references; construction of the documental tools (including a thesaurus); subsequent indexing of the references; and finally, the unification of all that tools in a global automated database which will allow the retrieval of the references according to the user's specifications. As an example, an informative database with all the information about written culture in the Quixote is exposed. It is concluded that this methodology allows to extract the information value (both implicit and explicit) from those indirect sources and serve it to the user quickly and efficiently.

Keywords: Written culture. Indirect sources. Informative databases. Information processing. Quixote. Thesaurus.

de tiempo y de recursos por parte del investigador.

Por ello, parte de la tarea del documentalista es ahorrarle ese tiempo y esfuerzo al investigador. Tradicionalmente, los bibliotecarios, archiveros y documentalistas se limitaban a tener las fuentes localizadas y organizadas para que el investigador pudiera buscar la información por su cuenta. Pero, como gestores de la información que somos, podemos ir más allá y proporcionar un servicio más detallado, ofreciendo a cada inves-

tigador, cada usuario, la información específica que necesiten dentro de cada fuente.

Por tanto, el presente trabajo de investigación intenta ser una aproximación al establecimiento de diversos criterios metodológicos que podrían emplearse para extraer el valor informativo relativo a una temática específica a partir de diversas fuentes de información, siendo nuestro objetivo principal la creación de diversas herramientas que condensen y optimicen ese valor informativo determinado y su agrupación en una macroherramienta que sea de utilidad para el usuario que necesite disponer de dicha información de manera organizada y el acceso rápido a ésta.

Dicha temática específica que hemos determinado para delimitar la información a extraer es la Cultura Escrita y la Historia del Libro y de la Lectura, temas que constituyen una de las líneas de investigación básicas para nuestro Grupo de Investigación (1). Sin embargo, la metodología que vamos a exponer a continuación es genérica y por tanto se puede extrapolar para el estudio de cualquier materia cultural-histórica.

Más allá de los eficaces métodos para seleccionar y localizar la información dentro de los documentos históricos tradicionales, nuestro grupo se ha especializado en investigar la extracción de la información a partir de fuentes menos visibles, que denominamos *fuentes* indirectas, las cuales no se suelen utilizar tradicionalmente en la investigación histórica o se utilizan en menor medida debido a que con ellas aumenta la dificultad que mencionábamos más arriba, esto es, su dispersión o la dificultad de localizar la información útil.

De todo ello, se desprenden los cometidos principales de esta investigación:

1. Crear una herramienta que sirva de base para el estudio de la Cultura Escrita, basándonos en una serie de fuentes menos habituales y utilizadas a través de la rigurosa revisión de la ésta.
2. Justificar la utilidad de la aplicación de las técnicas documentales al análisis de la Cultura Escrita. Esto es, aplicar distintas técnicas documentales de análisis de contenido a diversa documentación, desarrollando ciertas herramientas documentales que permitan entresacar el contenido relativo la materia.
3. Sentar bases documentales para la creación de un sistema de información dotado del mayor rigor y exhaustividad científica posibles para la investigación de aspectos relativos a la Cultura Escrita en otros marcos.
4. Aplicar lo expuesto al caso concreto del Quijote.

Para alcanzar estos objetivos, emplearemos diferentes recursos metodológicos, combinando técnicas de análisis de contenido, cuantitativas, documentales y de aplicación de las Nuevas Tecnologías de la Información.

2. Contextualización del objeto de investigación

2.1. La Cultura Escrita

2.1.1 Algunos apuntes sobre el concepto actual de la Cultura Escrita

Las técnicas de investigación que comentaremos se enmarcan en el paradigma o marco de estudio conocido como la Cultura Escrita, que es un campo difuso y complejo.

En efecto, es complejo porque no hay siquiera un consenso en la terminología por parte de distintas disciplinas y estudiosos. Se han sugerido conceptos nucleares como alfabetización, alfabetismo, Cultura Escrita, literacidad, literacia, etc.; y todos ellos suelen aportar matices propios pero también inconvenientes.

En este sentido, por un lado, el concepto anglosajón de *literacy* no se refiere sólo a la adquisición de la habilidad de leer y escribir, sino que involucra aspectos más amplios como los saberes, competencias, o el propio ambiente que conformaría lo que estudiosos como Chartier (2) han llamado la "cultura letrada" (Chartier, 1992). En todo caso, es difícil encontrar términos unívocos y universalmente aceptados, y de ahí la necesidad de expresiones o circunloquios como "prácticas alfabetizadoras" o "eventos letrados" entre otros. Del mismo modo, hay que disociar el concepto de leer y escribir de usos excluyentes, como la literatura. Así, escribir en el contexto de Cervantes es lo que realiza el escribano, mientras que, por ejemplo, para nosotros tiene una identificación sobre todo con el proceso creativo.

En segundo lugar, es un concepto difuso porque se acomoda al enfoque y los instrumentos de distintas disciplinas: el profesor, el sociólogo, el filólogo o el historiador ven distintos aspectos relevantes de la alfabetización.

De este modo, la Historia de la Lectura cobra relevancia para aclarar todos estos procesos, y en esta línea debemos citar algunas perspectivas recientes como la denominada "nuevos estudios de literacidad" y estudiosos de la Universidad de Lancaster. Así, por ejemplo, Google, la Feria de Libros de Frankfurt y la UNES-

CO han lanzado un proyecto de alfabetización para promover la lectura y la educación en todo el mundo. Google ofrecerá soporte a las instituciones participantes para que puedan contactar entre ellas y compartir programas y vídeos educativos y en Google Maps se podrán visualizar y ubicar todas las organizaciones que luchan contra el analfabetismo, útil a la hora de encontrar recursos para enseñar a leer y escribir. Incluye información relevante para profesores, facilitadores y trabajadores sociales: artículos, vídeos, libros, revistas, bitácoras, clubes, grupos de lectura, etc.

Vemos que se involucran aspectos no sólo educativos, sino relativos al futuro del libro, la diversidad de fuentes de información y de agentes alfabetizadores, los aspectos sociales, implícitos y explícitos, etc.

Esto es más obvio todavía en la versión anglosajona de este portal de alfabetización. Si en la versión española el material presentado se refiere sobre todo a alfabetización, los centros escolares y los educadores, en la anglosajona la consideración es mucho más amplia y da cabida a otros aspectos sociales. En ambos casos, el icono o emblema refleja muy bien este nuevo concepto de alfabetización, a caballo entre lo tradicional y lo moderno, el libro y lo digital, lo local y lo global, lo manuscrito y lo impreso e incluso hay una alusión clara al papel de la oralidad.

Como en la mayoría de conceptos científicos, hay cierta controversia sobre cómo delimitar la noción de Cultura Escrita. Aplicando un necesario criterio ecléctico, lo entendemos como un concepto omnicompreensivo, que abarca diversos planos. Esto se verá después en diversas investigaciones como nuestra línea de trabajo en torno a la Cultura Escrita dentro del Quijote.

Por ejemplo, involucra un plano psicopedagógico que reduce la Cultura Escrita a una habilidad o competencia aprendida. Sin embargo, la Historia Cultural y los nuevos estudios de literacidad (entre los que podemos citar los de Barton, 1994) nos enseñan que el alfabetismo no es un aprendizaje aislado o individual, sino una práctica social, de modo que los géneros y formas de la alfabetización no son fijas, sino que se construyen culturalmente. Incluso sucede que hay prácticas alfabetizadoras más visibles o con más prestigio que otras, pero sólo por una valoración social (por ejemplo, la literatura está mejor considerada que el cómic), y a su vez, dentro de cada género, hay un canon u obras dominantes.

La actualidad de estos estudios está constatando ahora que se ha pasado de una óptica muy

restrictiva, de la alfabetización como competencia o habilidad en que los únicos agentes eran el profesor y el alumno, a una visión mucho más amplia que subraya el papel de diferentes agentes y ámbitos: la escuela, pero también la familia, el barrio, el entorno urbano e incluso Internet. Algo que se puede constatar a través de las técnicas de estudio de la Cultura Escrita como hemos hecho en Cervantes.

2.1.2 Algunas líneas de estudio pertinentes de la Cultura Escrita

Desde hace aproximadamente medio siglo, la historia del libro se ha afianzado en la investigación internacional como una disciplina que abarca -al decir de los teóricos franceses esencialmente- varios enfoques diferentes y complementarios (3): a) por un lado, el enfoque economicista, que estudia el libro como objeto de comercio y sus repercusiones económicas; b) or otro, el estudio de la bibliografía material del libro relativa a los procesos de fabricación, la edición e impresión y todos sus aspectos materiales; y c) por último, el que consideramos el enfoque más interesante y que centra la mayor parte de nuestras investigaciones, que analiza al libro como conductor de texto y que tiene por objeto analizar la producción y la difusión de la cultura del libro.

Esta tercera perspectiva del “libro como contenido”, nos plantea cuestiones como los hábitos de lectura y escritura de la sociedad (marcándonos límites espacio-temporales más o menos extensos para delimitar el estudio), las vinculaciones de los autores con los diferentes estratos sociales y profesionales y las influencias que éstas tienen sobre el contenido de las obras creadas, e incluso se puede llegar hasta un prontuario para una sociología de la lectura.

Es dentro de este último aspecto, donde, a partir del estudio de las fuentes indirectas, intentamos alcanzar una panorámica general de *cómo y qué se leía*, determinando quiénes tenían la opción de leer —factor social: alfabetismo— quién tenía acceso a los libros —factor económico: precio de los libros— y, dentro de este círculo, quién acababa adquiriendo el hábito de la lectura —factor cultural—.

2.2. Las fuentes de información indirectas

Es fácilmente percible cómo el investigador tiene muchas opciones de partida como base para obtener información sobre múltiples y diferentes aspectos de la Cultura Escrita. Quizás, no obstante, para obtener una perspectiva global al respecto, fuera preciso que combinara la

información proveniente de las fuentes directas con las indirectas.

Entendemos como *fuentes directas* las que aluden claramente a la cuestión estudiada, esto es, aquéllas de carácter histórico, como la documentación archivística, jurídica y administrativa. Así, por ejemplo, tenemos protocolos notariales en la búsqueda de testamentos, almonedas, alguna correspondencia e inventarios *post mortem*, siempre consultados con las cautelas precisas. Asimismo, disponemos de instrumentos de control como los utilizados en las bibliotecas y en los archivos, que pueden darnos ciertos datos de cómo se percibían los libros y documentos a través de las diferentes épocas. Son, obviamente, las fuentes más utilizadas para el estudio de estos aspectos, por lo que están más procesadas y organizadas que el resto.

No obstante, tenemos otro tipo de fuentes que tocan tangencialmente el tema de la Cultura Escrita, y de forma no tan explícita como las primeras, sino de forma más imprecisa. Estas fuentes, que calificamos *fuentes indirectas*, tienen carácter histórico, literario y simbólico, y son de utilidad para especificar la visión que de los libros y de la lectura se tenía en un determinado momento histórico. Son múltiples, y van desde las fuentes tradicionales o narrativas —que incluyen memorias, diarios, epistolarios, y correspondencia; autobiografías, papeles propagandísticos, gacetas, avisos y noticias; o relatos y memorias de viajeros)—, a obras de carácter jurídico —como los procesos judiciales—, literario (obras literarias, representaciones teatrales y piezas liminares); o simbólico-plástico —literatura emblemática- (4).

Dichas fuentes son especialmente convenientes para el análisis de la visión —o percepción social— que de la Cultura Escrita se tenía en la época, dando multitud de interrogantes sobre su función, el uso cotidiano de los documentos o las prácticas de lectura. Todas estas fuentes pueden arrojar información crucial sobre los hábitos de lectura, escritura y culturales de la población en su época.

Por ello, y a pesar de que sí que se han realizado estudios sobre la Cultura Escrita a partir de ellas, no son tan numerosas o extensamente tratadas como las anteriores y creemos que aún ofrecen ricas posibilidades de estudio sobre esta rama del conocimiento. Muchas veces, para hacer una investigación histórica, la mayoría de investigadores se limitan a investigar en los documentos archivísticos tradicionales pasando por alto el valor informativo de este otro tipo de fuentes. Pero no debemos desdeñar el papel informativo que cumplen documentos que

originariamente no tenían esa intención de informar. De este modo, las obras literarias pueden aportar datos muy esclarecedores para muy diversos asuntos al respecto, si bien se hace necesaria una criba de los elementos literarios y ficticios (si los hubiera) de los netamente informativos para poder obtener una información lo más objetiva posible. Una aplicación correcta de las técnicas y herramientas documentales disponibles actualmente facilita esta labor y convierte estas obras originalmente no informativas en una fuente de información de alto valor.

2.3. Los Sistemas de Información Histórica (SIH)

Dado que ya hemos establecido que el principal objetivo de este trabajo gira en torno a la documentación aplicada a la investigación histórica del Libro y la Lectura, entendemos que es preceptivo estudiar del mismo modo el documento considerado como fuente de información. En este orden de cosas, los Sistemas de Información Histórica y Documental estudian y analizan los documentos para extraer información histórica. Por todo ello, parece preciso hacer una breve aproximación al concepto de dichos sistemas.

Es conveniente comenzar diciendo que un Sistema de Información Documental está constituido por un conjunto de subsistemas que se relacionan entre sí para lograr un objetivo común; si uno de los subsistemas falla, se altera el sistema total. Naturalmente un sistema computarizado proporciona mayor capacidad para el procesamiento de datos y la toma de decisiones.

En este mismo sentido, un Sistema de Información Documental Integrado permite un mayor acceso a la información dado que minimiza la redundancia de datos y facilita que todos los elementos, datos y registros se relacionen entre sí. Se asegura así el asequible flujo de la información, formulando políticas y fomentando el proceso de comunicación con los usuarios del Sistema. En definitiva, los Sistemas de Información Documental han asumido el papel central del manejo de la información de sus organizaciones y el diseño de sistemas de acceso.

Por su parte, los SIH, a juicio de García Marco (1995) nacieron como "una solución integradora: como un interfaz entre el tratamiento meramente archivístico y el análisis del dato histórico que ayude a optimizar el flujo de la información en el campo de las ciencias sociales y humanas", y su objetivo es resolver, de forma documental, las necesidades de los usuarios de información histórica. Esto es, los SIH resultan ser un ins-

trumento operativo que convierte los documentos históricos debidamente conservados en información, que previamente, tratada mediante los diferentes tratamientos documentales, es utilizada por el investigador para la producción de nueva información.

En este orden de cosas, el archivo histórico es un potencial centro de documentación y SIH, un sistema que a su vez labora por obtener de modo eficaz las notas de conservación, accesibilidad y difusión de la información histórica mediante operaciones de representación y organización de la información contenida en los documentos y que, finalmente, da a luz productos informativos propios como son las bases de datos históricas o facilita el acceso a bases de datos históricas producidas por otros sistemas de información.

En suma, los SIH se configuran como las herramientas más logradas en el campo que nos ocupa. Se hacen posibles merced al uso de las tecnologías informáticas e integran bases de datos y otros procedimientos similares que aseguren el correcto flujo de la información para la Historia. Desde las importantes reflexiones que García Marco (1995) ha dedicado a esta cuestión, procede enumerar las cuestiones de mayor interés relacionadas con los Sistemas de Información Histórica. Estos sistemas:

- a. Estructuran información en forma de bases de datos de carácter referencial, cuantitativo, textual, factual y multimedia.
- b. Evitan la duplicación de tareas de recopilación y organización documental en tanto en cuanto la información estructurada es accesible, conectable e incorporable a otros SIH.
- c. Propician la especialización de la documentación histórica por parte de las instituciones responsables de la construcción de tales sistemas.
- d. Contribuyen a normalizar la ciencia histórica y su crecimiento mediante la interacción de tres elementos: la propia fuente histórica, la historiografía y el trabajo del investigador
- e. Tratan de integrar todos los elementos referidos a las fuentes y a los resultados de las investigaciones, lo que incluye la normalización de las operaciones documentales y de la terminología, entre otros factores.

En consecuencia, un SIH no es una simple colección de fuentes o una descripción de unos determinados fondos, aunque lleven añadidas su digitalización y su etiquetado. Aunque dichas colecciones y fondos digitales constituyen componentes naturales de un SIH, éste es, ante todo, la integración en un sistema informático de

fuentes primarias y secundarias, datos y procedimientos de explotación bajo el paraguas de presupuestos teóricos y metodológicos compartidos por una comunidad de investigadores. De la misma manera, tampoco es un SIH la simple aplicación de herramientas informáticas a la investigación sobre el pasado. Un SIH es la aplicación sistemática de la informática a la investigación sobre el pasado, partiendo de instituciones y grupos de investigación comprometidos en compartir sus fuentes, datos y procedimientos, aunque, lógicamente, puedan sentir en el uso y conclusiones que saquen a partir del SIH.

3. Pautas metodológicas para procesar la información a partir de las fuentes indirectas.

3.1. Recursos metodológicos

Para alcanzar los objetivos previamente mencionados en la introducción, emplearemos y combinaremos diferentes recursos metodológicos. Veámoslos muy sucintamente:

1. *Técnicas documentales*, que nos permiten el estudio de las diversas obras como fuente de información. Dichas técnicas se fundamentan especialmente en el análisis de contenido de las fuentes, en lo relativo a sus aspectos conceptuales y cualitativos, para determinar el contenido informativo que aparece de forma implícita. Esto nos lleva a la creación de ciertas herramientas como un catálogo documental, unos índices toponomásticos y un tesoro *ad hoc* como lenguaje documental específico para la temática.
2. *Análisis crítico de la documentación*, lo cual nos permite tener una visión global de la evolución histórico-social de las materias objeto de tal corpus documental. Para ello, examinaremos las características generales e históricas de la temática.
3. *Técnicas filológicas* aplicadas a los textos seleccionados, que nos proporcionan una visión general de la Cultura Escrita dentro del universo de la historia, ofreciendo una cosmovisión cultural que será estudiada y analizada críticamente con la mayor profundidad posible.
4. *Técnicas cuantitativas*, basadas en la cuantificación de los conceptos relativos a la Cultura Escrita dentro de los fondos. Para ello, se analizan las coocurrencias de los términos de los diversos grupos conceptuales establecidos.

5. El empleo de las *nuevas tecnologías de la información*, que darán lugar a la confección un gestor documental automatizado aplicable a la temática en cuestión, y que permitirá la consulta y recuperación de la formación de forma pertinente y eficaz, así como permitir la plasmación de los conceptos en una representación cartográfica. En este sentido, elaboraremos una base de datos que deriva en catálogo de referencias, emplearemos aplicaciones informáticas para construir el tesauo, etc.

3.2. Etapas del trabajo

El procesamiento de la información a partir de estas fuentes indirectas conlleva un trabajo intelectual y de conceptualización, de localización y síntesis, que debe estar supeditado a los objetivos buscados. Sin embargo, de forma orientativa y genérica, puede estructurarse en fases como las siguientes:

3.2.1. Selección y procesamiento previo de las fuentes originales, a través de su análisis intuitivo y conceptual

Para ello, recopilaremos bibliografía general y especializada sobre el tema a tratar para su estudio (tanto de manera aislada como en conjunto), con objeto de extraer ideas y datos y obtener una panorámica global de dicho tema. Este conjunto bibliográfico será correctamente organizado y clasificado, para ser sometido después al proceso de extracción del contenido sobre Cultura Escrita.

Entre las tareas a realizar, tenemos:

1. Delimitación espacial, temporal y de profundización del tema sobre el que realizaremos la "criba" de información. Del mismo modo, efectuaremos acotaciones bibliográficas y limitaremos el volumen de la documentación empleada.
2. Selección de las fuentes indirectas que se vayan a utilizar para la búsqueda de información, tratando de seguir cierta diversidad dentro de éstos para poder tener una información contrastada.
3. Examen cuidadoso y comprensivo de las fuentes, intentando obtener las ideas o conceptos principales. Este proceso conceptual es uno de los más delicados y subjetivos de toda la cadena y depende casi enteramente de la capacidad de análisis y profundización del documentalista, que tiene que extraer, mentalmente, no sólo la información que aparece explícitamente en la fuente, sino

también la que se presenta de forma implícita.

3.2.2. Extracción de los fragmentos textuales o referencias

En esta etapa, (que puede fusionarse o no con la anterior, según el volumen de fondo a analizar y la dificultad y profundización en el tema), se deberá extraer cualquier referencia específica (es decir, fragmentos de texto o gráfico) relativa a algún aspecto sobre la Cultura Escrita dentro de los fondos, siendo el conjunto de referencias destinado a formar un catálogo de ellas que formarían parte de una futura base de datos con interrelaciones con las demás herramientas documentales.

Dichos fragmentos se irán recopilando en una base de datos y se irán indicando con términos tanto de materias como de nombres de autoridades (onomásticos) y lugares (topográficos), así como llamadas o reemisiones a la situación física del fragmento dentro de la obra.

Este proceso de extracción de la información se llevará a cabo, de forma intuitiva aunque metódica, y también con técnicas de extracción de contenido, que pueden ser automatizadas (como la extracción de términos recurrentes y fragmentos con alta presencia de dichos términos) o conceptuales como la indización manual o la elaboración de resúmenes por parte del documentalista.

Asimismo, y a la vez, también deberán extraerse los términos que condensan el contenido sobre la Cultura Escrita dentro de la obra y más concretamente dentro de las mencionadas referencias (los utilizados para indización por materias). Dichos términos acabarían convirtiéndose en la base para crear los descriptores de un tesauo que se utilizaría como herramienta metodológica.

En este sentido, las actividades a tratar podrían ser las siguientes:

1. Recopilación de referencias / textos sobre la Cultura Escrita dentro de los fondos.
2. Extracción de los unitérminos que servirán de base para los descriptores (tanto topomásticos como de materias).
3. Normalización de términos y recopilación de información sobre personajes, entidades, títulos de obras y lugares.
4. Cargado de referencias en una base de datos (la aplicación Access puede valer).

3.2.3. Construcción de las herramientas documentales específicas para ellos (incluyendo un microtesauro de descriptores) y posterior indización de las referencias.

A partir de toda esta información procesada en la fase anterior desarrollamos herramientas específicas para su indización, esta vez de carácter más riguroso y sistemático, así como para la recuperación de esta información. Estas herramientas documentales están interrelacionadas entre sí y gracias a ellas se multiplica el valor informativo de la fuente. Así, confeccionaremos las siguientes herramientas:

1) Una *base de datos* (que denominamos *catálogo*, aunque no en el sentido bibliotecario, sino en el sentido de listado) con una selección exhaustiva de todos los textos de la obra que hacen alusión a cualquier aspecto de la Cultura Escrita dentro de los fondos.

Tales textos deberán venir acompañados, además, de elementos que optimicen el catálogo y lo conviertan en una herramienta útil de búsqueda. Entre otros, sugerimos algunos de ellos: un número de registro, un breve resumen del contenido, su localización física dentro de la obra, diferentes descriptores normalizados, creados a partir de los términos extraídos en la etapa previa, sistematizados para obtener. Así, *descriptores toponomásticos*, tanto de autoridades de persona, entidad, título de obra, etc., como de lugares (los cuales integrarán los índices toponomásticos), o *descriptores de materias* (que después conformarán el tesauro)—, y cualquier otra información pertinente.

2) *Índices onomásticos y topográficos* donde se incluirán todas las autoridades, títulos de obras, entidades, lugares y, en general, nombres propios que se hallen en esa selección mencionada arriba. Estos descriptores onomásticos y topográficos provienen de los términos relativos a autoridades, entidades, títulos y lugares que habremos extraído de los fondos en la segunda fase, ya debidamente normalizados mediante la reglamentación pertinente con objeto de poder usarlos como forma de indización, y a la vez de recuperación de información.

Identificados en otras obras con los apéndices, tienen una gran utilidad para que el investigador pueda “moverse” rápidamente por la obra, ya que, al igual que la recuperación mediante descriptores de un tesauro facilita que se puedan hallar de forma eficiente todas las referencias a un mismo concepto de contenido, un índice onomástico también remite con rapidez a todas las referencias que en la obra se encuentren respecto a una persona, entidad, lugar, etc.

3) Y como tercera herramienta, confeccionaremos un *microtesauro* de descriptores *ad hoc*. Al igual que los índices toponomásticos tienen como objeto presentar los descriptores de autoridades, etc. de forma normalizada, el tesauro intenta configurarse como herramienta tanto de indización como de recuperación informativa, pero esta vez en cuanto a los descriptores de contenido temático.

Su elaboración se llevará a cabo de la forma recomendada, siguiendo las pautas de construcción de tesauros existentes (5), y partiendo tanto de los fragmentos como de los términos no normalizados (pero referentes al contenido) que seleccionamos y extraímos en las fases anteriores. Los diversos aspectos de la Cultura Escrita también se verían reflejados en este tesauro, siendo por ejemplo las categorías principales.

d. Por último, *integración de todo lo obtenido dentro de un gestor documental* automatizado que localice dichas referencias de acuerdo a las necesidades del usuario. Así, se crearía un instrumento informativo y documental completo, rápido, y eficaz, que sirva como fondo base, aunque procesado, de información relativa a la Cultura Escrita dentro de las fuentes originales que procesamos.

La idea es emplear los términos normalizados, tanto de los índices toponomásticos de autoridades como del tesauro de descriptores, para indizar los fragmentos (referencias) de texto que aludan a diversos aspectos de la Cultura Escrita. Asimismo, estos mismos términos son los que el usuario utilizaría a la hora de recuperar la información en forma de estos fragmentos textuales.

Con todo esto perseguimos la meta principal de tener la información preparada para ser servida al usuario investigador de acuerdo con sus necesidades y con rapidez, de forma que éste no tenga que invertir tiempo extrayendo el contenido que necesita de la fuente original, ya que toda la información sobre Cultura Escrita ya está procesada en una herramienta nueva.

4. Ejemplo / caso práctico: gestor documental e informativo sobre la Cultura Escrita en el Quijote

Como caso práctico de aplicación de estas pautas citamos una de las líneas de investigación de nuestro grupo que constituyó la base de una tesis doctoral (Martos, 2008). En este caso, se elaboró un Sistema de Información para el estudio de la Cultura Escrita dentro del Quijote de Cervantes, sobre aspectos tales como cultura

literaria, cultura popular, instituciones culturales, formulismos y referencias, lectura y escritura, documentos custodiados, etc.

En dicho trabajo se expuso que el Quijote, como obra literaria (y por tanto fuente indirecta), contiene una muestra representativa de la mayoría de las manifestaciones de la Cultura Escrita de la época, siendo un ejemplo singular de interacción entre documento escrito, oralidad y lectura.

Para la confección del mencionado Sistema de Información, extrajimos —tras su lectura y estudio sistemático— una cantidad aproximada de 750 fragmentos o referencias específicas sobre diversos aspectos de la Cultura Escrita dentro de la obra, elaborando un conjunto de herramientas como las que acabamos de describir e integrándolas dentro de un gestor automatizado y globalizado que serviría de ayuda y de base para cualquier investigador o usuario interesado en la temática.



Figura 1. Esquema de componentes del gestor de información

Este Sistema de Información o gestor documental, desarrollado a partir de la integración global de las diversas herramientas que hemos descrito, está constituido de la forma que se refleja en la figura 1 y cuenta con los siguientes componentes:

a. Una batería de referencias (alusiones a algún aspecto de la Cultura Escrita dentro del Quijote), obtenidas del Catálogo de referencias (1ª herramienta descrita al comienzo del capítulo), y adecuadamente indizadas. Cada registro / referencia tiene dos tipos de campos:

a.1. Campos informativos, como son el resumen del texto, el texto completo en sí o la localización física del texto dentro del libro y otros datos.

a.2. Campos descriptivos, con la descripción normalizada del contenido (no como en el resumen, donde dicha descripción es libre y orientativa) por los que se puede localizar la referencia dentro de la batería / catálogo: Campo onomástico de personas, de entidades, de títulos de obras, topográfico de lugares y de materias. Como se ha mencionado, estos campos estarán completados con los términos normali-

zados, de forma que la referencia quede adecuadamente indizada.

b. Una serie de documentos explicativos para que el usuario pueda seleccionar los términos normalizados al hacer las búsquedas avanzadas por campos (ver apartado siguiente). Estos documentos son: índice onomástico de personas, de entidades, de títulos de obras literarias y topográfico, u onomástico de lugares.

Se corresponderían, por un lado, con los informes descritos en la segunda herramienta, los

índices toponomásticos, Y, por otro lado, el índice conceptual del tesoro (en pdf).

c. Un menú buscador de referencias que ofrece dos opciones:

c.1. Una búsqueda por palabras en todos los campos. Esta búsqueda no tiene en cuenta las indizaciones hechas ni los términos normalizados de las herramientas.

c.2. Una búsqueda avanzada por campos, con los términos normalizados de los Índices Toponomásticos y el Tesoro.

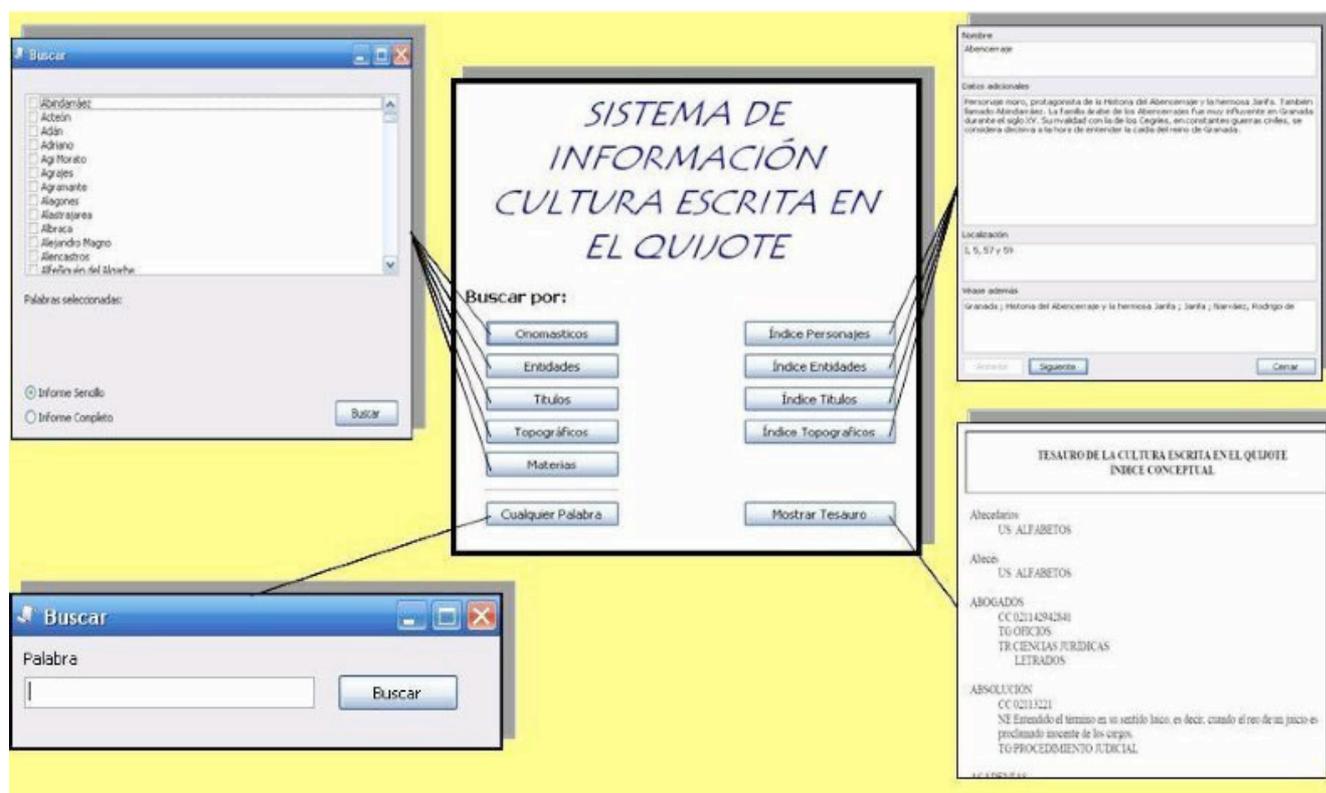


Figura 2. Presentación actual del gestor documental de la Cultura Escrita en el Quijote

El gestor sirve por ahora, principalmente, para hallar referencias relativas a la Cultura Escrita dentro del texto del Quijote, pero en el futuro se prevé también poder cargar otros trabajos relativos a la Cultura Escrita dentro del Quijote aunque no estén dentro de la propia obra (que también pueden ser halladas de igual forma que las referencias dentro del propio texto del Quijote), de forma que proporcionen un corpus bibliográfico de interés para cualquier investigación sobre la materia.

El usuario que quiera encontrar las alusiones en el texto del Quijote a cualquier aspecto de la

Cultura Escrita tiene dos opciones en el buscador:

- La primera es la más sencilla y simple: utilizar el buscador de texto completo, marcado por el rótulo Buscar cualquier palabra. Este buscador busca cualquier término en todos los campos. Esta búsqueda es más exhaustiva pero menos precisa.
- La otra opción es la más laboriosa pero la más precisa y eficaz. La búsqueda avanzada nos ofrece la posibilidad de buscar por campos, y en cada uno de ellos los términos normalizados, por lo que tenemos seguridad que

los resultados devueltos corresponden en gran proporción a la demanda de información que tenía el usuario.

El usuario se sitúa en cualquiera de los cinco campos (personajes, entidades, títulos, lugares o materias) y va seleccionando del menú desplegable aquéllos por los que quiere buscar. Para elegir los términos, es recomendable acceder a cada uno de los cinco documentos explicativos que hay en el menú de búsqueda, que corresponden, como ya comentamos, a los Índices de nombres propios.

La búsqueda se haría sobre la batería de referencias, arrojando los registros / referencias que cumplieran con los requisitos de búsqueda. Dichos registros hallados podrían aparecer, a elección del usuario, en modo simple (con sólo los campos informativos) o con el registro completo (con todos los campos, incluidos los de indización).

Pero este gestor es sólo el principio. El paso siguiente es el cargado de más fuentes bibliográficas sobre la materia, con su correspondiente indizado, para que puedan ser también recuperadas por la búsqueda.

Con todo ello, vemos que la construcción de una macroherramienta automatizada que emplee técnicas documentales de recuperación de información es capaz de proveer, de forma automática, toda la información textual dentro de la obra sobre cada uno de los conceptos relativos a la Cultura Escrita.

5. Conclusiones

A la luz de los resultados arrojados por este estudio y otros previos de temática similar, podemos inferir que, a partir de los criterios metodológicos expuestos, es posible extraer el valor informativo tanto explícito como implícito en estas fuentes indirectas y servirla al usuario de la forma más rápida y eficaz posible para él.

Por esto, consideramos que la aplicación de las técnicas documentales para el análisis de la Cultura Escrita tiene una gran utilidad, ya que, tras seguir las pautas sugeridas en este trabajo para el procesamiento de la información, se concluye en la creación de herramientas de análisis de contenido que nos permiten condensar el contenido relativo a este campo en todo tipo de documentos, incluso en las fuentes indirectas.

A partir de esto, comprobamos que las obras literarias son una buena fuente de información indirecta para el estudio de la Cultura Escrita. Sin embargo, seguramente la extracción de su

contenido (que, aparte de explícitamente, también existe de forma implícita en este tipo de fuentes) fuera más ardua y laboriosa si no se cuenta con herramientas de análisis de contenido adecuadas.

Con el uso de un sistema de información como el que hemos planteado, aumentan las posibilidades de que obras literarias como el Quijote u otras sean utilizadas como fuente de información para el estudio de la materia, con lo que se demuestra que la aplicación de herramientas documentales sobre fuentes indirectas como obras literarias (esto sirve para todo tipo de fuentes de información, pero vemos que en las indirectas ocurre con mayor intensidad) multiplica el valor informativo de la fuente como recurso para acceder a información de primera mano.

La creación de sistemas de información integrados facilita la exploración en profundidad de las materias, siendo de gran ayuda para todo aquel que pretenda investigar cualquier aspecto de la temática a tratar. Por tanto, se destaca las posibilidades de adaptación o implantación de estos sistemas como herramienta documental para estudiar los diversos campos científicos, humanísticos e históricos.

Asimismo, independientemente de la materia que se estudie y su marco temporal y local, es interesante que los investigadores conozcan las ventajas reales que tiene la implantación de un tesoro idóneo y ya desarrollado dentro de éste, tanto considerando cada sistema de información aisladamente como el conjunto de los sistemas españoles que podrían relacionarse a través de este instrumento de control terminológico.

Notas

- (1) Grupo de Investigación ARDOPA (Archivística y Documentación del Patrimonio). Vid.: ardo-pa.unex.es.
- (2) Es fundamental para nuestro trabajo la obra de R. Chartier. Vid. entre otros muchos títulos para nuestro cometido aquí: Chartier, R. (1985b), (1987a), (1987b), (1988), o (1995). Al decir de Ricardo García Cárcel, Roger Chartier es, sin duda, el historiador que más ha influido en la proyección de la cultura en los últimos diez años. Tras un inicial interés en los años setenta por la problemática de la historia de la educación, desenvolviéndose en sus primeros escauceos metodológicos bajo la sombra de la Escuela de Annales irrumpe en el horizonte historiográfico de modo bien visible en los años ochenta, a través de la dirección del III volumen de la Historia de la Vida Privada (Chartier, 1985a) y sobre todo de la codirección, junto a H.J. Martín, de la Histoire de l'edition française -3 vols., 1982, 1984 y 1985, respectivamente. Para este apartado seguimos a Le Goff, J.; Chartier, R.; y Revel, J. (1998).

- (3) Dichos enfoques al estudio de la Cultura Escrita y la historia del libro y la lectura, fueron mencionados con mayor detalle en nuestro artículo, presentado para la anterior edición de Ibersid, Martos García, A. y Vivas Moreno, A. (2007).
- (4) El tema de las fuentes indirectas ya ha sido tratado también por miembros del grupo en anteriores ediciones de Ibersid, como por ejemplo: Vivas Moreno, Martos García et al. (2005), (2006) o el ya aludido Martos García y Vivas Moreno (2007).
- (5) La construcción de tesauros y las pautas para ello son temas extensamente tratados. Quienquiera que desee tener una cierta metodología para la construcción de tesauros, puede acudir a las normas ISO 5964:1985 "Documentation -- Guidelines for the establishment and development of multilingual thesauri" y 2788:1986 "Documentation -- Guidelines for the establishment and development of monolingual thesauri", que en español se corresponden con la norma UNE 50125:1997 "Directrices para el establecimiento y desarrollo de tesauros multilingües" y la norma UNE 50106:1990 "Directrices para el establecimiento y desarrollo de tesauros monolingües"; o bien a las obras de expertos como Lancaster o Currás.

Martos García, A.; Vivas Moreno, A. (2007). Bases para un sistema de información para el estudio de la Historia de la Lectura en Extremadura. // Ibersid: revista de sistemas de información y documentación = journal of information and documentation Systems. 1 (2007) 165-176.

Vivas Moreno, A.; Martos García, A. [et al.] (2005). La visión de los archivos en el Antiguo Régimen hispano a través de las Memorias y relatos de viajeros: Análisis de su representación y sistematización metodológica. // Scire: Representación y organización del conocimiento. 11:1 (2005) 47-64.

Vivas Moreno, A., Martos García, A.; [et al.] (2006). La visión de los archivos y la cultura escrita en el Antiguo Régimen. // Documentación de las Ciencias de la Información. 29 (2006) 237-258.

Referencias

- Barton, D. (1994). *Literacy: An Introduction to the Ecology of Written Language*. Oxford y Cambridge, MA: Blackwell, 1994.
- Chartier, R. (dir.) (1985a). *Historia de la vida privada*, tomo 3, *Del Renacimiento a la Ilustración*. Madrid: Taurus, 1985.
- Chartier, R. (1985b). *Pratiques de la Lecture*. Marsella: Rivages, 1985.
- Chartier, R. (1987a). *Lectures et lecteurs dans la France d'Ancien Regime*. París, Seuil, 1987.
- Chartier, R. (1987b). *Les usages de l'imprimé*. París: Fayard, 1987.
- Chartier, R. (1988). *Cultural History Between Practiques and Representations*. Cambridge: Un., 1988.
- Chartier, R. (1992). *El mundo como representación*. Barcelona: Gedisa, 1992.
- Chartier, R. (1995). *Livres bleus et lectures populaires. Mi XVIIe siècle-debut XIXe*. // *Culturas populares: diferencias, divergencias, conflictos, en Europa entre los siglos XVI y XVII*. Barcelona: Gedisa, 1995.
- Documentación. *Directrices para el establecimiento y desarrollo de tesauros monolingües: UNE 50106-90 (ISO 2788:1986)*. En: *Documentación*. Tomo 2: normas fundamentales. Madrid: AENOR, 1994. 171-217.
- Documentación. *Directrices para la creación y desarrollo de tesauros multilingües. ISO 5964:1985 (UNE 50125:1997)*.
- García Marco, F. J. (1995). *Los sistemas de información histórica: entre la Archivística y la Historia*. // Ruiz Rodríguez, A. (Ed.). *Manual de Archivística*. Madrid: Síntesis, 1995. 103-184.
- Le Goff, J.; Chartier, R.; Revel, J. (1998). *Libro*. // *Diccionario del saber moderno. La Nueva Historia*. Bilbao: Ed. Mensajero, 1988. 91 y ss.
- Martos García, A. (2008). *Creación de un sistema de información sobre La Cultura Escrita en el Quijote* (Tesis doctoral dirigida por A. Vivas Moreno). En prensa.