

Sistema de información histórica para la colección de papeles varios del Archivo Histórico de la Universidad de Salamanca

Agustín Vivas Moreno

Facultad Biblioteconomía y Documentación
Universidad de Extremadura

0.1. Resumen

Se presenta el sistema de información histórica diseñado para la Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca, su proceso de desarrollo y las conclusiones metodológicas y teóricas obtenidas. El trabajo partió del análisis del proceso informativo documental, estableciendo la recuperación y la difusión de la información como sus fines últimos. Seguidamente, se realizó un análisis documental e histórico riguroso de la colección que resultó en su descripción mediante las técnicas archivísticas. En tercer lugar, se construyó el sistema de indización para la búsqueda y recuperación de la información, basado en la utilización de lenguajes documentales, en concreto de un tesaurus diseñado a la medida. Además del sistema automatizado, la investigación aportó una metodología novedosa para el análisis y tratamiento archivístico de las Colecciones Misceláneas y se constató a través de ella la utilidad de los gestores de bases de datos, los lenguajes de indización documental y los tesauros como instrumentos de información archivísticos, y, en general, las importantes conexiones que entroncan la Archivística con las Ciencias de la Documentación.

Palabras clave: Sistemas de información histórica. Colecciones misceláneas. Descripción archivística. Tesauros. Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca.

0.2. Abstract

The process of development of a historical information system designed for the Various Papers Collection of the University of Salamanca Historical Archive is presented, and its methodological and theoretical conclusions addressed. First, the informative process was established, considering information retrieval and dissemination as its final goals. Thereafter, a rigorous historical and documentary

analysis of the collection was done, which resulted in its careful description by means of archival techniques and standards. Thirdly, the indexing system was built, based in the use of a thesaurus, developed specifically for this collection. An innovative methodology for the treatment of miscellaneous archival collections was created. In this process, the advantages of data base management systems, indexing languages and thesauri for archival information were demonstrated, and, in general, the decisive connections of Archival Science with Information Science.

Keywords: Historical information systems. Miscellaneous collections. Thesauri. Archival description. University of Salamanca. Historical archives. Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca.

1. Introducción

1.1. Objetivos básicos de los Sistemas de Información Histórica

Nuestra ponencia pretende analizar el documento como objeto de análisis común a la Archivística y a la Historia, esto es, como fuente de información histórica, objeto de estudio de los llamados Sistemas de Información Histórica (García Marco, 1994,1995; López Yepes, 1997, 1998). Por otro lado, somos conscientes de que el aumento de las necesidades informativas en el tiempo presente se traduce en el progresivo uso de las tecnologías. En consecuencia, la estrecha relación entre Informática e Historia ha diseñado técnicamente auténticos Sistemas de Información Histórica. En definitiva, Archivística, Historia y Nuevas Tecnologías serán las bases metodológicas en las que estarán basadas las páginas que siguen que tienen como objeto el esbozo de un Sistema de Información Histórica para la Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca.

Podemos resumir, antes de adentrarnos en el núcleo de la exposición, y aunque sea de forma breve, los tres objetivos fundamentales que, a mi entender y al de otros estudiosos, procuran los Sistemas de Información Histórica:

1. Integrar el análisis archivístico y la investigación histórica sin por ello prescindir de la organización del documento y de las aportaciones que realizan las Ciencias de la Documentación (1).
2. Relacionar el mensaje que portan los documentos y la creación de conocimientos partiendo de los datos contenidos en las fuentes con el control y preservación de las unidades de instalación.
3. Actuar de puente entre el contenido informacional de los documentos primarios localizados y las necesidades informacionales y las representaciones cognitivas de los usuarios. En consecuencia, organizar los datos

históricos de cara a su análisis cualitativo y cuantitativo tratando de confeccionar representaciones documentales para la recuperación como los descriptores controlados.

En definitiva, los Sistemas de Información Histórica favorecen la creación de productos informativos propios —por ejemplo, las bases de datos históricas— y aseguran el correcto flujo de información para la Historia. Así, las técnicas documentarias cumplen sus objetivos: se hacen operativas al aplicarse a un campo del saber determinado, y se ponen al servicio de la obtención de nuevos conocimientos científicos (López Yepes, 1998, p. 100).

1.2. Características de la colección

Se trata de una sección facticia o colección documental, esto es, un conjunto de documentos reunidos según criterios no objetivos, y que por tanto, no conservan una estructura orgánica ni responden al principio de procedencia archivístico, según el cual no se debe alterar la organización dada al fondo por la unidad productora, y los documentos generados por una entidad o institución no deben mezclarse con los de otras corporaciones. En consecuencia, se trata de un conjunto de documentos reunidos de forma facticia por motivos de conservación debido a su especial interés histórico.

Suelen ser características de estas colecciones facticias la procedencia múltiple y difusa, la miscelaneidad temática, la amplitud cronológica, la diversidad regional y la vastedad documental. Igualmente, es particularidad de estas secciones la conjunta presencia de documentación archivística (testimonio material de un hecho o acto realizado en el ejercicio de sus funciones por personas físicas o jurídicas) y de material librario que responde a una documentación que, por sus características, debiera estar ubicada en una biblioteca. Todo ello origina que, de forma usual, la documentación localizada en estas colecciones presente una especial importancia histórica, que hace de ellas reductos muy útiles para historiadores especializados en determinadas temáticas.

Esta Colección ha sido poco trabajada. Al margen de lo que un día fue mi tesis doctoral y algunos artículos que he realizado al respecto, no conocemos nada publicado, ni tan siquiera de forma indirecta e incompleta (Vivas, 1994, 2000a, 2000b, 2000c, 2000d). Asimismo, la Colección ha sido mínimamente utilizada como base de otras investigaciones históricas. Por otra parte, ha estado hasta el momento fuera de la programación y planificación descriptiva del Archivo, hasta tal punto, que no ha recibido ningún tratamiento archivístico.

A continuación, se proporcionan, casi de forma telegráfica, una serie de características principales de la Colección objeto de nuestro Sistema de Información Histórica:

1. La colección documental está compuesta por 279 unidades de instalación, de las cuales se encuentran disponibles 206.
2. Hay fundamentalmente dos tipos de unidades de instalación: el volumen encuadernado (69,42%) y el legajo (30,58%). Los volúmenes, en la mayoría de los casos, aparecen encuadernados en pergamino, con una botonadura de cuerdas para abrochar. En el lomo, en ocasiones, dos o tres remaches de cuero sujetos con fibras entrecruzadas. En las hojas de guardas suelen utilizarse papel procedente de documentos o libros impresos, y, en pocas ocasiones, pergamino. El tamaño es variado, aunque de modo general puede estimarse en unas dimensiones de 200x300mm. Los legajos, por su parte, son atados de documentos protegidos por tapas de cartón. También de tamaño variable, predominando los de 210x310mm.
3. El número de agrupaciones documentales es de 4.602, lo que supone una media de 22,34 documentos por unidad de instalación.
4. En lo concerniente al análisis de los documentos por su condición de ser manuscritos o impresos se da un predominio de los documentos impresos (63,34%) frente a los manuscritos (36,66%), a pesar de la cuantía de estos últimos.
5. En lo referente al análisis cronológico, la Colección reúne documentación desde el XVI hasta el XIX. En este sentido, se caracteriza por un predominio de los documentos fechados en el XVIII (61,37%), seguido de los del XVII (31,15%). Al margen, encontramos documentos de los siglos XIX (4,05%) y XVI (3,43%). El centro neurálgico cronológico lo encontramos en la primera mitad del XVIII, aminorándose la cantidad de documentos en la medida en que nos alejamos de este período.
6. En lo que respecta al análisis topográfico, nuestra miscelánea destaca por la presencia de documentación datada en múltiples lugares de fuera de la Península, lo que define a la colección como cosmopolita, universal y alejada del localismo provinciano. Las principales zonas topográficas son Madrid, Salamanca y Roma, seguidas de lejos por Lisboa, Sevilla, Valladolid y Granada. Si entre las tres primeras, y como prueba de concentración topográfica, suman cerca de un 60 % del total de los documentos reconocidos, el otro 40% se caracteriza por una destacada dispersión topográfica.
7. En lo referente al análisis idiomático, predominan los documentos en castellano, seguido del latín. Marginalmente encontramos documentos en portugués, francés, italiano, e incluso algunos en catalán, griego y chino.
8. La Colección se caracteriza por tener heterogéneas y múltiples procedencias documentales. En este orden de cosas, la procedencia mayorita-

ria es la conventual (aproximadamente el 64% de las unidades de instalación), destacando mayoritariamente la jesuítica (58,01%) y la dominica (38,93%). Las unidades de instalación provenientes del Colegio Real del Espíritu Santo de Salamanca pasaron a formar parte del A.U.S.A. cuando tuvo lugar la expulsión de la Compañía de Jesús de España en 1767. Sobresale el citado fondo personal del P. Pedro Abarca (S.I.), o las obras de Luis de Losada (S.I.). Por su parte, las unidades procedentes del Convento de San Esteban de Salamanca pasaron a formar parte del A.U.S.A. con la exlaustración de la órdenes regulares del XIX, destacando el fondo del P. José Barrio (O.P.). Además, también contamos con unidades procedentes de otras Órdenes religiosas: Franciscanos (1,54%), Cistercienses (0,76%), o Agustinos Calzados (0,76%).

En lo que respecta a las procedencias no conventuales, hay que destacar, en primer lugar, el papel destacado de los fondos colegiales. Resalta el Colegio Mayor de Cuenca (60%) —que fundamentalmente se determina en memoriales de pleitos, conflictos jurisdiccionales varios localizados en las diócesis de Jaén, Cuenca y Guadalajara, y sermones y oraciones fúnebres y panegíricas varias pronunciados por colegiales o dedicadas a ellos— y, en menor medida, el Colegio Mayor de Santiago el Zebedeo (15%). Por otra parte, merece especial mención el fondo procedente del archivo personal de Lorenzo Velasco González (20%), personaje local influyente, con documentación de finales del XVIII y principios del XIX. Por último, también hay algunos documentos procedentes del Obispado de Salamanca (5%).

9. Múltiples y diversas tipologías documentales, esto es, los documentos derivan de heterogéneas actividades y competencias. Contamos con documentación constitutiva, probatoria e informativa. De este modo, están presentes, entre otras, Reales Cédulas, alegaciones jurídicas, cartas privadas, libros de cuentas, bulas, breves, fueros, constituciones, etc.
10. Por contenidos, los Papeles Varios son especialmente útiles, por su temática específica, para el historiador especialista en Órdenes Religiosas y en Universidades. Los temas jesuíticos y los dominicos son abundantes y llenos de detalles. Por su parte, son analizables multitud de temas vinculados a la Universidad de Salamanca.

Una segunda categoría la componen lo que hemos dado en llamar *temas prioritarios o preferentes*, esto es, temáticas que, sin ser específicas de la colección, destacan sobremanera por su copiosa documentación. En este sentido, son de especial interés histórico los sermones y oraciones fúnebres y panegíricas del XVIII, la documentación historiográfica sobre la

secesión de Portugal), la sátira y sus vertientes políticas, sociales y religiosas, la abundante documentación sobre aspectos económicos y sociales de la ciudad de Salamanca y su partido judicial en el XIX, el copioso material para el estudio de la fiesta en los siglos XVII y XVIII como manifestación de la religiosidad popular, y la documentación sobre Colegios y otras instituciones universitarias.

Y en tercer y último lugar, lo que venimos denominamos como *temas complementarios*, esto es, contenidos relevantes en nuestra Colección y que podrían completar ciertas investigaciones históricas. Destacan determinados asuntos internacionales de carácter fundamentalmente político, cuestiones de política nacional principalmente de los reinados de Felipe V, Fernando VI y Carlos III, asuntos de política eclesiástica, conflictos jurídicos y pleitos por causas de derecho privado y sucesiones, aspectos de religiosidad y mentalidades, historia de minorías, temática inquisitorial, contenidos económicos, genealogías de familias y linajes, poesía, pastorales diocesanas y abundante material librario que, por su condición, debieran estar en la Biblioteca.

2. El Sistema de Información Histórica

Una vez explicado el concepto de Sistema de Información Histórica, y analizado las características de la Colección, estamos en condiciones de explicar el transcurso de nuestra investigación. Para ello, analizaremos los objetivos, la metodología, los resultados y las fases del trabajo.

2.1. Objetivos

Nuestro principal objetivo puede definirse como la realización de un análisis descriptivo e histórico de la Colección con objeto de llevar a cabo la recuperación y la difusión de la información allí contenida. En este orden de cosas, varias fueron las pretensiones en este empeño:

- a) *Analizar documental e históricamente* la Colección, esto es, partiendo del documento, principio esencial del quehacer documental, extraer las características generales de la colección e interpretarlas sistemáticamente.
- b) Realizar *instrumentos de descripción* archivísticos para la Colección, tarea primordial, pues está encaminada a poner los documentos en servicio, es decir, hacer de ellos un útil disponible para la sociedad y la investigación histórica. Por consiguiente, consistirá en revelar el contenido y el carácter de los documentos, y facilitar su localización.
- c) Confeccionar un *tesauro "ad hoc"* como lenguaje documental para la Colección. Se trata, por consiguiente, de trazar una fórmula que sirva para

la búsqueda y recuperación de la información por parte de los usuarios, consistente en sintetizar y representar el contenido de los documentos.

- d) Y diseñar un instrumento que permitiera buscar y recuperar la información de forma automatizada, esto es poder gestionar y manipular los datos que conformen nuestro instrumento de descripción, así como permitir búsquedas de documentos específicos para recuperar la información demandada de forma pertinente

En definitiva, se pretendía crear un proceso informativo-documental para llevar a cabo la recuperación y la difusión de la información contenida en la Colección, de tal forma que produjera un nuevo estado de conocimiento en el receptor, mediante el cual éste la adquiriera con el propósito de producir nueva información. Entiendo, por tanto, el proceso informativo-documental como la creación, puesta en forma y comunicación de mensajes con una finalidad informativa.

2.2. Metodologías y resultados

Para conseguir estos grandes objetivos se han empleado unos determinados recursos metodológicos, que han dado lugar a unos resultados.

- a) Uno de ellos ha sido la utilización de *técnicas de descripción archivísticas*. Ha constado de dos etapas:
 1. Análisis de las diferentes unidades de instalación con objeto de realizar un examen de contenido de las diferentes agrupaciones documentales y facilitar la localización de dicha documentación.
 2. Análisis exhaustivo de cada una de las piezas documentales (documentos sueltos) y de las unidades archivísticas (expedientes), tanto en sus caracteres internos como externos. El resultado son, por un lado, las tablas y relaciones de contenido como instrumento de control archivístico, y, por el otro, el catálogo como instrumento de información.
- b) Un segundo recurso metodológico ha sido el empleo de los *lenguajes documentales* como instrumento para la búsqueda y recuperación de la información. Con tres etapas en esta ocasión:
 1. Examen de los documentos y determinación de su contenido.
 2. Identificación y selección de los conceptos principales extraídos de documentos mediante análisis intelectual y transformación posterior en términos de indización.
 3. Selección de términos de indización que expresaran los contenidos de los documentos y establecimiento de relaciones jerárquicas, asociativas y de equivalencia. El resultado de esta metodología son los índices de persona, entidad, lugar y materia, y el tesauro.

- c) El uso de la *tecnología de la información* es otra de las metodologías empleadas. Igualmente consta de varias etapas:
 1. Diseño de la estructura de los datos siguiendo el modelo relacional.
 2. Implementación mediante el gestor de bases de datos dBASE V.
 3. Incorporación de herramientas para gestionar y manipular los datos que conforman el catálogo, elegir descriptores del tesoro y navegar por ellos, y realizar búsquedas documentales por medio de descriptores y en texto libre.

El resultado es el gestor documental automatizado para la Colección de Papeles Varios (Vivas Moreno, 2000e)

- d) Y, por último, el análisis histórico de la documentación, que consiste en el examen de la documentación desde el punto de vista del investigador interesado por la historia. De este modo, he realizado diversos análisis sectoriales a los documentos ubicados en la Colección: formal-descriptivo, tipográfico, cronológico, de procedencias, topográfico, idiomático, y de contenido.

El resultado se manifiesta en la confección de las características generales para la Colección, producto sintético como consecuencia del trabajo analítico anterior.

En definitiva, Archivística, Historia y Nuevas tecnologías componen conjuntamente las líneas evolutivas sobre las que he basado metodológicamente el trabajo de investigación, con tres características:

1. El análisis del documento desde una visión interdisciplinar, esto es, como “concepto-cruce de la Historia y la Archivística”.
2. La aplicación de técnicas de indización documental a instrumentos de descripción archivísticos.
3. Y la contribución de la tecnología informativa, como base de elaboración de los llamados Sistemas de Información Histórica.

2.3. Fases de trabajo

Por otra parte, estos recursos metodológicos empleados para cada uno de los objetivos se han aplicado siguiendo una determinada secuenciación.

- a) En un primer momento, efectué una *primera toma de contacto* con la Colección en dos niveles:
 1. *Análisis del estado de la cuestión*: Se rastreó la bibliografía, y se examinó lo argumentado en la literatura científica para el tratamiento archivístico de colecciones misceláneas.

2. Acercamiento inicial a las unidades de instalación y a las agrupaciones documentales ubicadas en la Colección.
- b) La segunda fase del trabajo está dedicada, por tanto, al *análisis descriptivo e histórico* de las unidades de instalación. En ella se ha empleado la mayor cantidad de recursos. Para su desarrollo se han realizado tres grandes cometidos:
1. *Análisis exhaustivo de las unidades de instalación y de cada uno de los documentos* que se localizan en dichas unidades. Labor ardua y fatigosa, de cuño positivista, pero inexistente hasta hoy, y, por tanto, imprescindible, dado que el documento se configura como el fundamento de nuestro quehacer documental.
 2. *Creación de un proceso informativo documental para llevar a cabo la recuperación y la difusión de la información*, como fines últimos de la actividad documental (López Yepes, 1978, 1981, 1995). Pretendía que la información documental presentada fuera adquirida por el receptor con el propósito de producir nueva información, esto es, se configurara como un proceso informativo documental con dos características: que aportara algo potencialmente informativo, y que contribuyera a crear información que contiene a su vez información.

Para su cumplimiento se trabajó en tres niveles de forma simultánea:

- La confección de los diversos *instrumentos de descripción*.
- La elaboración de un *lenguaje de indización* propio, esto es, un sistema que facilitara la representación formalizada del contenido de los documentos para permitir la búsqueda y recuperación de la información solicitada por los usuarios. Además, para los descriptores de materias confeccioné una lista estructurada de conceptos, y establecí con ellos relaciones de equivalencia, jerárquicas y asociativas, de tal forma que los términos de indización aislados quedaron convertidos en verdaderas familias de descriptores que conforman un Tesoro ad hoc, destinado a representar de manera unívoca el contenido de los documentos y de las consultas.
- El diseño de un *gestor documental automatizado*.

Todo ello se insertó en un CD ROM en el que se puede recuperar y difundir la información requerida.

3. Etapa de *síntesis* y extracción de características generales. En esta fase comenzamos a interrogar el material recogido desde diversas hipótesis y cuestionamientos, y extraímos conclusiones referidas al volumen documental, la tipografía, la cronología, las zonas de pro-

ducción, los idiomas, las procedencias, los contenidos y la importancia para la investigación histórica.

- d) Y, por último, la tercera fase de trabajo, estuvo destinada a la síntesis de las conclusiones. Constituyen, por consiguiente, la hipótesis ya demostrada, las respuestas a los problemas planteados.

3. El tesoro, catálogo y gestor documental automatizado: Bases del Sistema de Información Histórica

Tesoro, Catálogo y Gestor Documental Automatizado son las bases fundamentales de nuestro Sistema de Información Histórica. A continuación se analizan su diseño y características.

3.1. Tesoro

El tesoro es uno de los componentes fundamentales para la operatividad del Sistema de Información Histórica confeccionado para la Colección, dado que es un instrumento que sirve de puente entre el vocabulario utilizado por el indizador de los documentos y el usuario de los mismos. (2)

Sus funciones habían de ser las siguientes: en primer lugar, la normalización del vocabulario, esto es, la eliminación de la ambigüedad del lenguaje en nuestra disciplina; en segundo lugar, la inducción, pues se invita al usuario potencial a la consulta de temas colaterales, gracias al establecimiento de una serie de referencias cruzadas que indican las relaciones asociativas, jerárquicas y preferenciales que se pueden producir entre la terminología que lo compone; y en tercer lugar, la representación o la correspondencia que se debe establecer entre los descriptores y los conceptos presentes en los documentos.

Por otro lado, nuestro instrumento ofrecía ciertas ventajas que han sido tenidas en cuenta: es un lenguaje controlado; es un lenguaje combinatorio postcoordinado; reúne además de los descriptores, los no-descriptores o equivalencias semánticas; recoge relaciones de jerarquía y asociación, que facilitan la localización de los descriptores más apropiados cuando se realiza la indización de los documentos y, sobre todo, en el momento de la formulación de las consultas; reúne la mayor univocidad semántica, evitando la sinonimia y la polisemia; puede actualizarse y ponerse al día periódicamente; y permite la incorporación de notas de alcance históricas, que ayudan sustancialmente a explicar y describir el sentido y significado de los descriptores.

En modo alguno ambicionamos reunir todo la terminología utilizada en nuestra disciplina para ponerla a prueba mediante su uso sistemático en la indización de documentos de distintos fondos y usos. Este tesoro ha sido creado específica y especialmente para dicha sección facticia. Evidentemente, esto con-

lleva características singulares fruto de la propia idiosincrasia de la sección, que se explican a continuación:

1. Es un tesoro *a posteriori*, es decir, hemos utilizado los propios documentos a indizar entre las fuentes empleadas. Así pues, el presente tesoro muestra la ventaja de que la selección de los descriptores está oportunamente pensada para los documentos que posteriormente se indizarán.
2. Es un tesoro *misceláneo*, resultado de la propia idiosincrasia de la Sección, para la que está pensado. La realización de un sistema de información documental que aglutine, debe considerar entre sus obstáculos el grave problema de la amplitud temática —Estado y gobierno, derecho y justicia, hacienda, ejército y guerra, religión, educación y cultura, economía, población y sociedad—, cronológica —desde el XVI hasta el XIX— y documental —abundantes tipologías documentales y copioso material librario—. Todo ello ha originado graves problemas de normalización terminológica, que se han visto acrecentados por el carácter histórico de los términos: la ortografía, la evolución lingüística del castellano, el olvido y cambio de significado de los términos o la diversidad regional son algunos problemas específicos con los que nos hemos encontrado.

Ante esta situación nuestro tesoro presenta dos características que pudieran parecer contradictorias: por un lado, su rigidez, dado que no hemos pretendido confeccionar un tesoro exhaustivo y hemos optado por un modelo esquemático cuyo nivel de profundidad y extensión del campo terminológico no es completo; y por otro, su flexibilidad, procurando hacer de nuestro lenguaje documental un instrumento versátil, función que se consigue a través de unas relaciones asociativas elaboradas y exhaustivas.

3. Es un tesoro *descifrado*. Uno de los mayores problemas que presenta un tesoro histórico es el establecimiento del significado concreto de los descriptores y sus relaciones semánticas. Dominar la disciplina histórica se hace indispensable, no sólo para comprender la ingente flexibilidad relacional existente entre los descriptores, sino también para inferir y conocer su verdadero alcance. Para solventar problemas de este tipo realizamos notas de alcance histórica para cada descriptor, de tal forma que el usuario potencial sabe exactamente no sólo el significado histórico del término, sino también el sentido del mismo y su exacta aplicación en el tesoro que presentamos.

Expuestas las anteriores características, se presentan en la tabla 1 las áreas temáticas y el número de descriptores y no descriptores que lo componen.

Respecto a la evaluación del tesoro, podemos extraer los siguientes datos:

- La *ratio de precoordinación* —es decir, la que procede de dividir el número de veces que aparecen las palabras que constituyen los términos del tesaurus por el número de descriptores— es de 2,34. Es una ratio alta, fruto de la especificidad de los documentos que presentamos y de la concreción de los títulos de los campos semánticos.
- La *tasa de equivalencia* —o sea, la ratio entre el número de descriptores y no descriptores— es de 1,07. Presenta, por tanto, una tasa elevada (gran cantidad de no descriptores) que mejora la coherencia y la precisión de la indización, ya que tiene como resultado el aumentar el número de entradas alfabéticas del tesaurus y, en consecuencia, la probabilidad de que un término en lenguaje natural que aparece dentro de un documento o de una consulta coincida con una entrada del tesaurus.
- La *tasa de enriquecimiento* —es decir, la ratio entre el número de relaciones jerárquicas y asociativas, por una parte, y el número de descriptores por otra— es de 2,9. Se trata, por consiguiente, de una ratio elevada. Esto tiene como ventaja lo estructurado y entrelazado que está el tesaurus, puesto que el usuario encontrará abundantes reenvíos de unos descriptores a otros; pero presenta, quizá, el inconveniente de que, al haber tantas relaciones, no resulte fácil examinarlas.
- Por último, la *tasa de flexibilidad* —esto es, la proporción de palabras significativas simples utilizadas para constituir los descriptores compuestos (es decir, precoordinados) y que se encuentran así mismo como descripto-

Áreas temáticas	Descriptores principales	Sinónimos	Total descriptores	%
01 Estado	232	142	374	21,25
02 Derecho y Justicia	110	154	264	15
03 Hacienda	71	71	142	8,07
04 Fuerzas Armadas y Guerra	49	49	98	5,57
05 Iglesia y Religión	204	192	396	22,5
06 Educación y Cultura	100	75	175	9,94
07 Economía	74	74	148	8,41
08 Población y Sociedad	55	73	128	7,27
09 Ciencias y Documentos Auxiliares	18	17	35	1,99
TOTAL	913	847	1760	100

Tabla 1. Áreas temáticas y número de descriptores y no descriptores del tesaurus

res no compuestos o como no descriptores— es mediana: alrededor de 0,33. Esto se debe esencialmente por una parte, al número de palabras que están presentes en el tesoro al formar parte del nombre de una institución, y por otra, a los términos, en ocasiones demasiado específicos, que son necesarios para representar conceptos históricos.

3.2. Catálogo

La confección de un catálogo como base de nuestro sistema de información documental nace de las características intrínsecas de la Colección. De la imposibilidad de llevar a cabo instrumentos macrodescriptivos, de la diversa y variada procedencia de nuestra documentación, del alto valor histórico que consideramos que tiene el conjunto documental (Tornel, 1991), de la presencia de material librario en no poca cantidad, de la falta de organicidad en nuestra colección, y de la necesidad de recurrir prioritariamente a tareas microdescriptivas, se desprende y justifica plenamente la confección de un catálogo, esto es, de un análisis descriptivo de cada pieza documental ubicada en nuestro objeto de trabajo. De esta manera, el contenido de las diferentes agrupaciones documentales podrá ser buscado, localizado y recuperado por medio de un instrumento de información que se encontrará también informatizado.

Los elementos descriptivos que componen nuestra ficha se estructuran y distribuyen en áreas o campos con denominación propia, siguiendo las ISAD (G):

- Área de *mención de identidad*, donde se incluye la información esencial para identificar la unidad que se describe: signatura, fecha y lugar de producción.
- Área de *contexto*, donde se incluye información acerca de la producción, organización y materias de la unidad de descripción (3).
- Área de *contenido*, referida fundamentalmente a un resumen de contenido: nota-resumen de alcance y contenido
- Área de *acceso y utilización*, donde se incluye lo referido a la lengua y escritura de los documentos y las características materiales: manifestación manuscrita/impresa (4), localización, lengua de los documentos, características físicas (formato, tamaño, conservación, ilustraciones, soporte, etc.)
- Área de *notas*: donde se hará constar la estructura del documento y cualquier otro tipo de observaciones.

La ficha concluye con el *incipit* y los descriptores onomásticos y de entidad, lugares y materias.

Las áreas y los elementos incluidos en las ISAD(G) no mencionados son base del estudio de las características generales válidas para toda la colección: el título de la colección (Papeles Varios), el nivel de descripción (Colección) y el

volumen de la unidad de descripción (279 unidades de instalación) en el Área de mención de identidad; los nombres de los productores (Jesuitas y Dominicos fundamentalmente), la historia administrativa (múltiple, en ocasiones desconocida por tratarse de numerosos productores), las fechas de acumulación de la unidad de descripción (variadas, dependiendo del productor), la historia de la custodia y la forma de ingreso (esencialmente en el XVIII y XIX tras la expulsión de los Jesuitas y la exclaustación de los Dominicos) en el Área de contexto; la información sobre valoración, selección y eliminación (al tratarse de documentación histórica, de valor permanente), los nuevos ingresos (quizás motivado por la presencia de volúmenes de Papeles varios en la Biblioteca Universitaria) y el sistema de organización (no existe) en el Área de contenido; y las condiciones de accesibilidad (las típicas para documentación histórica) en el Área de condiciones e acceso y utilización. Por su parte, el Área de materiales relacionados, que consiste en dar noticias sobre la localización de originales, existencia de copias, unidades de descripción relacionadas o fuentes complementarias en el propio archivo, documentación complementaria en otros archivos o instituciones y notas de publicación o bibliografía ha sido tratada con búsquedas exhaustiva en la Biblioteca Universitaria de Salamanca y en la ingente obra de Simón Díaz (1983) (5) para el XVII, la de Aguilar Piñal (1981) para el XVIII, y en obras de referencia especializadas para determinadas Órdenes Religiosas (8).

En conclusión, nuestra ficha catalográfica intentará describir los documentos en tres sentidos: en primer lugar, en cuanto a su contenido y localización temporal, resumiendo la información contenida en pocas líneas, a fin de orientar al consultante sobre la conveniencia o no de su lectura, y dotándole de una datación lo más completa posible; en segundo lugar, en sus características formales —volumen, tamaño, soporte, conservación, tipología, etc.—; y en tercer lugar, en su localización espacial dentro del fondo al que pertenece y de la unidad de instalación en la que está inserto, a través de la signatura y la localización respectivamente (Cruz, 1995, p. 283-287).

En definitiva, el catálogo que presentamos de manera impresa y automatizada debe entenderse como un instrumento archivístico de consulta, que forma parte del modelo global que denominamos cadena documental, íntimamente relacionado con un lenguaje de indización, confeccionado de manera complementaria, como es el tesoro. De este modo, las técnicas documentarias son aplicables al campo archivístico, teniendo como principal fin la recuperación y la difusión, objetos últimos de la actividad documental.

3.3. Diseño del gestor documental

Como ya hemos comentado, el propósito del sistema confeccionado es la gestión y manipulación de los datos que conforman el catálogo, así como permitir las

búsquedas de documentos específicos para recuperar la información demandada de forma pertinente. En este orden de cosas, este breve apartado tendrá por objeto explicar brevemente las características internas y en herramientas que conforman nuestro gestor documental (Ministerio..., 1995 ; Pressman, 1995) (6).

Para el diseño de la estructura de datos se ha seguido el modelo relacional. En dicha estructura hay dos tablas principales: la de documentos y la de descriptores, que han sido relacionadas con una cardinalidad N:M, esto es, varios documentos pueden estar relacionados con varios descriptores y viceversa. Se ha seguido, pues, el modelo entidad/relación. Ello ha sido implementado siguiendo el diseño de gestión de datos dBASE V.

La estructura de la tabla de documentos es la que aparece en la tabla 2. Como puede observarse, se han utilizado principalmente campos "memo". A pesar de tener el inconveniente de no permitir indizarlos, presentan la ventaja de consentir una longitud indefinida, lo que resulta de gran utilidad dadas las peculiaridades de los documentos que describimos. Aquellos atributos que son de otro tipo están indizados, dando lugar a índices de campo completo y no de palabras clave.

Nombre	Tipo de datos	Longitud	Índice
Clave doc	Númérico	10	s
Signatura	Carácter	70	n
Comienzo	Carácter	85	s
Lugarfecha	Memo	=	n
Ms imp	Memo	=	n
Conservación	Memo	=	n
Ilustración	Memo	=	n
Lenguaje	Memo	=	n
Localización	Memo	=	n
Tamaño	Memo	=	n
Otros	Memo	=	n
Título	Memo	=	n
Ref	Memo	=	n

Tabla 2., Estructura de la tabla de documentos

Nombre	Tipo de datos	Longitud	Índice
Clave_des	Numérico	10	s
Tipo	Carácter	1	n
Descrip	Carácter	85	s
Tesaurus	Carácter	20	s

Tabla 3. Estructura de la tabla de descriptores

Por su parte la tabla de descriptores presenta la estructura que aparece en la tabla 3. Se encuentra, por consiguiente, indizada por su clave, por la clave correspondiente del tesaurus y por el alfabético del descriptor, permitiendo así su recuperación y ordenación por el valor numérico de la clave, por el alfabético del descriptor o a través del tesaurus.

Los descriptores —de persona/entidad, lugar y materia— se encuentran relacionados con los documentos. Ello da lugar a que se puedan realizar búsquedas entre los descriptores y seleccionar aquellos documentos con los que están relacionados, tanto a través de operadores lógicos, como por medio de la inclusión de los registros que contengan descriptores que formen parte de la estructura jerárquica del tesaurus.

Por otra parte, el sistema incorpora una herramienta para seleccionar descriptores que permite navegar por ellos utilizando su clave, el texto inicial, el tipo de descriptor, así como las relaciones del tesaurus en el caso de los de materia.

Por último, se ha diseñado una tabla que hemos denominado “alfa”, cuya estructura se presenta en la tabla 4. Esta tabla se utiliza a modo de fichero invertido. Su generación es automática mediante una opción del menú que opera en tres fases:

- *Extracción de términos*: se eliminan las palabras de una longitud menor o igual que dos.

Nombre	Tipo de datos	Longitud	Índice
Término	Carácter	15	s
Campo	Carácter	1	n
Clave_doc	Numérico	4	n
Idf	Numérico	15	n

Tabla 4. Características de la tabla alfa

- *Cálculo de pesos idf*: para la recuperación posterior, se calculan las frecuencias documentales de cada término y en base a ello se realiza la ponderación *Inverse Document Frequency* (Salton, 1983; Salton, 1989).
- *Normalización de los vectores documentales*: se utiliza para ello la norma euclídea, esto es, la división de los componentes por cada módulo. Todo ello conlleva que sea menor el cálculo a realizar durante la recuperación.

En el momento de la recuperación se puede introducir una cadena de doscientos caracteres como máximo, que es vectorizada utilizando los mismos métodos que para los documentos. Tras ello, se halla la similitud entre cada documento y la pregunta (7).

El sistema permite recuperar los 101 documentos con mayor similitud. Asigna una pertinencia de 100 al que mayor similitud tiene y luego va decreciendo proporcionalmente. Es necesario destacar también que la búsqueda no se hace de manera exacta, sino por truncamiento. Es decir, si en la petición se incluye el término “a”, se buscarán todos aquellos términos que comiencen por “a” pertenecientes a los documentos, lo cual, evidentemente, ralentizará la recuperación. En definitiva, el propósito del sistema es la gestión y manipulación de los datos así como la posibilidad de hacer búsquedas, no sólo por descriptores —lo cual no comporta excesiva dificultad—, sino también en texto libre. Sin embargo, no tiene posibilidades de impresión por sí mismo, sino que genera ficheros que pueden ser manejados como datos de fusión mediante un procesador de textos (en nuestro caso Word Perfect), permitiendo así modificar el texto antes de imprimirlo, y lanzar la impresión desde el mismo.

4. Conclusiones

Además de los resultados científicos y técnicos alcanzados, es menester destacar las conclusiones de carácter metodológico:

1. La utilidad de la aplicación de lenguajes de indización documental y tesauros a instrumentos de información archivísticos; esto es, la adaptación de las técnicas documentales a la disciplina archivística.
2. El método de tratamiento archivístico de las Colecciones misceláneas pasa por la descripción de las unidades de instalación y de las piezas documentales y unidades archivísticas. Fruto de dicho tratamiento son las relaciones de contenido (instrumento de control) y el catálogo (instrumento de información).
3. El uso de la tecnología de la información es apropiado para que se consume el proceso informativo documental, es decir, para que se produzcan las fases de difusión y recepción de la información.

4. Nuestro distema de información histórica se basa sustancialmente en un lenguaje documental estructurado, un instrumento de descripción por-menorizado y un eficaz tratamiento automatizado que tiene por objeto la recuperación y la difusión de la información.

5. Notas

- (1) “Ser una solución integradora: una especie de interface entre el tratamiento meramente archivístico y el meticoloso análisis del dato histórico que ayude a optimizar el flujo de la información en el campo de las ciencias sociales y humanas.” (Cfr. García Marco, 1995, p. 105).
- (2) “Un tesoro es una lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario en la indización de los documentos y de las consultas” (Asociación Francesa para la Normalización, 1981).
- (3) Tanto la forma de ingreso como el nombre del productor y su historia se exponen de forma global, al estudiar las procedencias y las características históricas de la Colección.
- (4) Dada la presencia abundante de material librario era necesario un elemento que recogiera este aspecto, a pesar de que no viniera recogido en las ISAD(G).
- (5) El problema de esta obra es que está sin acabar, llegando solo hasta la letra “P”
- (6) Para este apartado, véase Vivas Moreno y Guerrero Bote (2000). Agradezco desde aquí al Prof. Dr. Vicente Guerrero sus numerosas ayudas y útiles consejos.
- (7) Para ello, se utiliza el coeficiente del coseno. Estas funciones han sido objeto de estudio durante mucho tiempo. Sin embargo, algunos estudios vienen a demostrar que hay un grupo de medidas entre las que no existen diferencias apreciables. Entre ellas una de las más favorecidas es la de coseno, que se puede expresar también como el producto escalar de los vectores normalizados. En cierta forma es lógico, ya que, fruto de la normalización, se eliminan las diferencias existentes entre los vectores de mayor y menor longitud, que corresponderían respectivamente a los documentos que tienen más y menos términos.

6. Bibliografía

- Aguilar Piñal, F. (1981). Bibliografía de autores españoles del s. XVIII. Instituto Miguel de Cervantes C.S.I.C., 1981-1995.
- Association française de normalisation (1981). Règles d'établissement des thésaurus monolingues. Paris, AFNOR, 1981. (AFNOR Z 47-100).
- Bibliotheca carmelitana, notis criticis et dissertationibus illustrata, 1927.
- Bibliotheca Universa Franciscana, suve alumnorum trium ordinum S.P.N. Francisci, qui ab ordine seraphico condito, usque ad praesentem diem, Latina, sive alia quaris lingua scripto aliquid consignerunt, Enciclopedia uvilloti athenaeo et syllabo uradingiano... Fr. Joannis de Soto, Matritis: ex Typographia Causae V. Matris de Agreda, anno 1732.
- Cruz Mundet, J. R. (1995). Manual de Archivística. Madrid : Fundación Germán Sánchez Ruipérez, 1995.

- Eug de Uriarte, J. (1904). Catálogo razonado de obras anónimas y seudónimos de autores de la Compañía de Jesús, pertenecientes a la antigua asistencia española: con un apéndice de otras de los mismos, dignos de especial estudio bibliográfico. Madrid : Sucesores de Ribadeneyra, 1904. 6 vols.
- Eug de Uriarte, J. ; Lecina, M. (1929-1930). Biblioteca de escritores de la Compañía de Jesús pertenecientes a la antigua asistencia de España, desde sus orígenes hasta el año 1773 . Madrid : Imprenta Gráfica Universal, 1929-1930. 2 vols.
- François, Jean (1961). Bibliothéque générale des écrivains de l'Ordine de Saint Benoît. Louvain : Hêrerlé, 1961. 4 vols.
- García Marco, F. J.(1995). Los sistemas de información histórica : entre la Archivística y la Historia. // Ruiz Rodríguez. A. (Coord.). Manual de Archivística. Madrid : Síntesis, 1995. 103-184.
- García Marco, F. J. (1994). Sistema de información en historia regional : Los mudéjares del Jalón y el Jiloca medios en el siglo XV. Zaragoza : Universidad de Zaragoza, 1994.
- Kaeppli, T. O. P. (1970) Scriptorum Ordinis Praedicatorum Medii. Romae : S. Sabinae, 1970. 3 vols.
- López Yepes, J. (1978). Nuevos estudios de documentación: el proceso documental en las ciencias de la documentación social. Madrid: Instituto Nacional de Publicidad, 1978.
- López Yepes, J. (1981). El estudio de la Documentación: metodología y bibliografía fundamental. Madrid: Tecnos, 1981.
- López Yepes, J. (1995). La documentación como disciplina: Teoría e historia. Pamplona : Eunsa, 1995.
- López Yepes, J. (1997). Los caminos de la información : cómo buscar, seleccionar y organizar las fuentes de nuestra documentación personal. Madrid : Fragua, 1997.
- López Yepes, J. (1998). Las bases de datos históricas. // Anales de Documentación. Murcia : Universidad de Murcia. 1 (1998) 99-125.
- Ministerio para las Administraciones Públicas (1995). METRICA, Versión 2.1. Madrid : Tecnos, 1995.
- Placer, G., O. de M. (1968). Bibliografía mercedaria. Madrid: Revista Estudios, 1968. 2 vols.
- Pressman, R. S. (1995). Ingeniería del Software: un enfoque práctico. London : McGraw-Hill, 1995 (trad. 3ª ed.).
- Salton, G. ; Buckley, C. (1988). Term weighting approaches in automatic text retrieval. // Information Processing and Management. 24 (1988) 513-523.
- Salton, G. ; McGill, M. J.(1983). Introduction to modern information retrieval. New York : McGraw-Hill, 1983.
- Santiago Vela, Gregorio de (1913). Ensayo de una biblioteca ibero-americana de la Orden de San Agustín. Madrid : Imprenta del Asilo de Huérfanos del Sagrado Corazón de Jesús, 1913, 8 vols.
- Simón Díaz, J. (1975). Jesuitas de los siglos XVI y XVII : Escritos localizados. Madrid : Universidad Pontificia de Salamanca y Fundación Univ. Española, 1975.
- Simón Díaz, J. (1977). Dominicos de los siglos XVI y XVII: escritos localizados. Madrid : Universidad Pontificia de Salamanca y Fundación Univ. Española, 1977.

- Simón Díaz, J. (1983-1999). *Bibliografía de la literatura hispánica*. 3ª ed. Madrid : Instituto Miguel de Cervantes C.S.I.C, 1983-1999.
- Sommervogel, Carlos, S. I. (1890). *Bibliothèque de la Compagnie de Jesús*. Bruxelles-París, 1890. 12 vols.
- Tornel Corbacho, C. (1991). El inventario en el panorama archivístico español . // *Irargi*. 4 (1991) 297-325.
- Vivas Moreno, A. (1994). *Historiografía de la Secesión de Portugal en los fondos documentales (Mss. y P.V.) de la Biblioteca Universitaria de Salamanca: Una primera aproximación al tema*. // Carabias Torres, A. Mª. (ed.). *Las relaciones entre Portugal y Castilla en la época de los descubrimientos y la expansión colonial*. Salamanca : Univ. de Salamanca, 1994
- Vivas Moreno, A. (2000a). *La colección documental de Papeles Varios del Archivo Histórico de la Universidad de Salamanca*. // Rodríguez-San Pedro Bezares, L.E. (Coord.). *Actas del V Congreso Internacional de Universidades Hispánicas*. Salamanca : Servicio de Publicaciones, Universidad de Salamanca ; Junta de Castilla y León, , 2000. XXX.
- Vivas Moreno, A. (2000b). *La Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca. Análisis de la documentación : procedencias, características e importancia histórica*. // *Brocar : Cuadernos de Investigación Histórica*, 24 (2000).
- Vivas Moreno, A. (2000c). *Procedencias documentales de la Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca*. // *Bilduma*. (2000).
- Vivas Moreno, A. (2000d). *La Colección de Papeles Varios en su contexto*. // Rodríguez-San Pedro Bezares, L.E. (Coord.) *Nueva Historia de la Universidad de Salamanca*. Salamanca : Universidad de Salamanca, 2000.
- Vivas Moreno, A. ; Guerrero Bote, V. (2000). *Gestor documental automatizado para la Colección de Papeles Varios del Archivo Histórico de la Universidad de Salamanca*. // *Cuadernos de Documentación Multimedia*. Madrid : Univ. Complutense, 2000.
- Vuaddingus, Lucas, Fr. (1650). *Scriptores ordinis minorum quibus accessit syllabus illorum, qui ex eodem Ordine pro fide Christi fortiter accuberunt*. Roma, Anno 1650. Reed.: Minerva G.M.B.H. Frankfurt, 1967.