

Una aproximación al concepto de recuperación de información en el marco de la ciencia de la documentación

José Antonio Salvador Oliván *
Rosario Arquero Avilés **

Artículo recibido:
7 de junio de 2005.
Artículo aceptado:
8 de junio de 2005.

RESUMEN

La naturaleza interdisciplinar de la recuperación de información plantea problemas no fáciles de resolver a la hora de establecer una definición válida para todas las ciencias relacionadas con este campo. El presente artículo pretende ofrecer una definición de recuperación de información en el contexto de la ciencia de la documentación. Para ello, se presenta en primer lugar una visión global a través de los principales conceptos proporcionados en la literatura del campo y se analiza su influencia y relación interdisciplinar con la informática, la ciencia de la información y la documentación. Posteriormente, se realiza un análisis intradisciplinar dentro de la propia documentación con el objetivo de

* Universidad de Zaragoza, España. jaso@posta.unizar.es

** Universidad Complutense de Madrid, España. carquero@ccinf.ucm.es

establecer sus propias fronteras y determinar su propia identidad.

Palabras clave: Recuperación de información; Relaciones con otras ciencias.

ABSTRACT

An approach to the concept of information retrieval in library and information science

José Antonio Salvador Oliván and Rosario Arquero Avilés

The interdisciplinary nature of information retrieval poses difficulties not easy to solve when trying to establish a sound definition for all the related sciences of the field. This paper proposes a definition of information retrieval in the context of library and information science and thereby offers a global vision taking the main concepts given in the literature of the field. It also examines the influence of these concepts and their interdisciplinary relationship with informatics, and library and information science. Subsequently, an intradisciplinary analysis is made within library and information science in order to establish its own frontiers and determine its own identity.

Keywords: Information retrieval; Relationship with other fields.

INTRODUCCIÓN: CONCEPTO GLOBAL Y CARACTERÍSTICAS GENERALES DE LA RECUPERACIÓN DE INFORMACIÓN

Cuando Calvin N. Mooers introdujo por primera vez en 1950 el término *information retrieval* en la literatura de documentación, lo definió como «la búsqueda de información en un stock de documentos, efectuada a partir de la especificación de un tema».¹ Sólo un año más tarde, el mismo autor ampliaba esta definición al manifestar que

1 C.N. Mooers, "The theory of digital handling of non-numerical information and its implications to machine economics", en *Technical Bulletin* No. 48. Cambridge, MA: Zator Co., 1950 (Ponencia presentada en Association for Computing Machinery, Rutgers Univ., New Brunswick, NJ, 1950, March 29).

la recuperación de información abarca los aspectos intelectuales de la descripción de información y su especificación para la búsqueda, y también cualquier sistema, técnica o máquina que se utilice para llevar a cabo la operación.²

De esta manera, quedaban definidos los tres problemas que esta actividad debía de tratar:

- ¿Cómo definir y organizar la información?
- ¿Cómo especificar la búsqueda?
- ¿Qué sistemas y técnicas utilizar para estos procesos?

Un campo de estudio se define por los problemas que trata y los métodos que utiliza para resolverlos,³ y estos problemas, que no son mutuamente exclusivos, reflejan a grandes rasgos las principales y diferentes áreas de trabajo, investigación y discusión que existen en el campo de la recuperación de información.

Jean Tague-Sutcliffe identifica las siguientes variables o elementos que definen el campo de la recuperación de información:⁴ (a) colección de documentos o bases de datos, (b) representación de la información, (c) usuarios, (d) consultas y frases u oraciones de búsqueda-estrategias de búsqueda, (e) intermediarios de búsqueda, (f) proceso de búsqueda, y (g) evaluación de la recuperación. Hay que tener en cuenta que estas variables están situadas a un macro nivel y que en cada una de ellas se pueden estudiar diferentes aspectos. No todos los aspectos tienen la misma importancia para todos los profesionales e investigadores; algunos se centran más en la entrada en el sistema (representación y almacenamiento de la información), otros en la salida (búsqueda), y otros en el propio sistema (diseño de Sistemas de Recuperación de Información [SRI]). Los diferentes intereses, derivados de los problemas que hay que resolver ponen de manifiesto un hecho conocido, y es que no existe una única idea sobre lo que significa investigar o trabajar en la recuperación de información, aunque, como dice Blair, esta falta de coherencia no es inusual en un campo incipiente.⁵

Esta falta de coherencia puede venir determinada por la naturaleza interdisciplinar de la recuperación de información (RI). La ciencia de la información,

2 C.N. Mooers, "Zatocoding applied to mechanical organization of knowledge", en *American Documentation*, 1951, vol. 2, pp. 20-32.

3 T. Saracevic, "Information science: origin, evolution and relations", en Perri Vakkari and Blaise Cronin (Eds), *Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives*. London: Taylor Graham, 1992, pp. 5-27.

4 J. Tague-Sutcliffe, "The pragmatics of information retrieval experimentation, revisited", en *Information Processing and Management*, 1992, vol. 28, n° 4, pp. 467-490.

5 D.C. Blair, *Language and representation in information retrieval*, Amsterdam: Elsevier Science Publishers, 1990.

la informática, la documentación, la psicología, la lingüística, la lógica, la inteligencia artificial, etcétera, son algunas de las disciplinas que en mayor o menor medida contribuyen a la investigación teórica y aplicada de la recuperación de información.

A pesar de que en estos poco más de 50 años se han producido cambios importantes en las tecnologías, soportes y métodos empleados, así como en el tipo y formatos de información, los problemas identificados por Mooers siguen siendo la base de la recuperación de información, cuyo principal objetivo, y en esto hay un consenso generalizado, es proporcionarle información relevante al usuario para satisfacer su necesidad de información.

La consecución de este objetivo es el tema central de la recuperación de información. El cómo lograrlo ha atraído el interés de investigadores de diversas disciplinas y provocado una actividad multi e interdisciplinar en las tres direcciones señaladas anteriormente, conformando a la vez tres características esenciales y que consideramos el *leit motif* de la evolución y existencia del campo de la recuperación de información:

- Interdisciplinariedad
- Estrecha relación con la tecnología de la información
- Participación activa en la evolución de la sociedad de la información

Estas tres características son compartidas por otros campos y ciencias como la documentación, la ciencia de la información y la informática. La recuperación de información, como componente importante de estas tres ciencias, es partícipe de estas características y constituye el marco para comprender el pasado, presente y futuro de este campo y las cuestiones y problemas a los que se enfrenta.

OBJETIVOS

El hecho de ser un campo interdisciplinar favorece por un lado la presencia de múltiples conceptos, pero por otro complica y dificulta el hecho de proporcionar un único concepto que sirva para todas las ciencias, y crear un panorama complejo y difícil de sintetizar.

El propósito de este trabajo es conceptualizar y articular una definición de la recuperación de información en el marco de la ciencia de la documentación, limitando al mismo tiempo su campo de aplicación y objetos de estudio, algo nada fácil frente a la intersección de las diferentes ciencias mencionadas anteriormente e incluso otras subdisciplinas propias de la documentación.

METODOLOGÍA

La metodología seguida para conseguir el objetivo final ha sido la siguiente:

- Analizar los diversos conceptos de recuperación de información propuestos por autoridades reconocidas en este campo.
- Establecer su relación con la ciencia de la documentación, la ciencia de la información y la informática.⁶ La última, por ser consustancial a la recuperación de información, las dos primeras porque constituyen el marco esencial en el que se desarrolla nuestra actividad profesional y docente, y aunque para diversos autores constituyen una misma ciencia, existen también quienes piensan que, aunque están íntimamente relacionadas, son independientes. Sin entrar a valorar si la ciencia de la información (*information science*) y la documentación son la misma ciencia, en este trabajo realizaremos un análisis separado de ambas ya que pensamos que puede enriquecer notablemente la visión global de la recuperación de información, y nos va a permitir ir delimitando su zona de influencia y los problemas que intenta resolver cada una de ellas, a la vez manifestar ciertas diferencias entre la ciencia de la documentación en España y la ciencia de la información en Estados Unidos, al menos en lo que se refiere al campo de la recuperación de información.
- Determinar su identidad dentro de la ciencia de la documentación, y establecer los límites con otras subdisciplinas.
- Por último, y como síntesis de todo lo anterior, proponer una definición de recuperación de información en el marco de la ciencia de la documentación.

ANÁLISIS DEL CONCEPTO EN LA LITERATURA

En este apartado se realiza un análisis del concepto en la literatura para comprobar si existe un consenso generalizado sobre el significado de recuperación de información y si se ha producido alguna variación con respecto a la definición dada originalmente por Mooers. Los conceptos seleccionados que se presentan a continuación proceden en su gran mayoría de publicaciones

6 Bajo el término ciencia de la documentación, y según la definición propuesta por José López Yepes, incluimos tanto la biblioteconomía, como la archivística y documentación. (J. Lopez Yepes, *La documentación como disciplina. Teoría e historia. Segunda edición actualizada y ampliada*. Pamplona: EUNSA, 1995, pp. 321-322).

consideradas como libros de texto en las asignaturas en cuyo título figura “Information Retrieval”, y que son impartidas en las escuelas de biblioteconomía y ciencia de la información de Estados Unidos; libros que han sido escritos por autoridades reconocidas en este campo.

Van Rijsbergen dice que

[...]la recuperación de información está relacionada con la recuperación de aquellos documentos que sean probablemente relevantes para la necesidad de información del usuario expresada en una petición.⁷

En su libro,⁸ todo un clásico, destaca tres áreas principales de investigación en recuperación de información: análisis de contenido, estructuras de información y evaluación.

Gerard Salton dice que «[...] la recuperación de información está relacionada con la representación, almacenamiento, organización y acceso a ítems de información».⁹ Idéntico concepto proporciona Ricardo Baeza-Yates, señalando que

[...]el área de recuperación de información ha crecido mucho más allá de sus objetivos primarios de indización de texto y búsqueda de documentos útiles en una colección.¹⁰

Charles Meadow comienza el prefacio de su libro:

La recuperación de información es un proceso de comunicación. Es un medio por el que los usuarios de un sistema o servicio de información pueden encontrar los documentos, registros, imágenes gráficas, o registros de sonido que satisfagan sus necesidades o intereses.¹¹

Más adelante continúa:

Esencialmente, y de una manera simple, la recuperación de información implica encontrar información deseada en un almacén de información o base de datos. El

7 C.J. Van Rijsbergen, “A non-classical logic for information retrieval”, en Sparck Jones, K.; Willett, P. (Eds.) *Readings in Information Retrieval*, San Francisco: Morgan Kaufman Publishers, Inc., 1997, pp. 268-272.

8 C.J. Van Rijsbergen, *Information retrieval*. 2nd edition. London: Butterworth, 1979, p. 5 <<http://www.dcs.gla.ac.uk/Keith/pdf/>

9 G. Salton; M. J. McGill, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983, p. 1.

10 R. Baeza-Yates; B. Ribeiro-Neto, *Modern Information Retrieval*. New York: ACM Press, 1999, pp. 1-2.

11 C.T. Meadow; B. R. Boyce; D. H. Kraft, *Text Information Retrieval Systems*. Second Edition. San Diego, California: Academic Press, 1999, Prefacio, xv., p. 2.

proceso de buscar selectivamente información en una base de datos puede verse que arranca desde dos puntos diferentes. Un punto de comienzo es el usuario de la información. [...]En el otro punto de inicio, una persona o agencia toma la decisión de recoger, organizar y almacenar la información para posteriormente buscarla y recuperarla.

Para Frederick W. Lancaster,

[...]recuperación de información, tal y como se utiliza habitualmente, es sinónimo de búsqueda de literatura; es el proceso de buscar en una colección de documentos (utilizando el término documento en su más amplio sentido) para identificar aquellos que tratan de un determinado tema.¹²

Más recientemente, manifiesta que

[...]en la actualidad la recuperación de información “convencional” significa la búsqueda *online* en bases de datos electrónicas, de forma interactiva y en tiempo real. Normalmente, esto implica que el usuario construye una estrategia de búsqueda usando términos con distintas relaciones lógicas (booleanas) y que el programa de búsqueda simplemente divide la base de datos en dos conjuntos: elementos recuperados y elementos no recuperados.¹³

Considera sistemas avanzados aquellos que se desvían de esta norma y que permiten plantear la petición de búsqueda de manera diferente a la booleana y/o que dan lugar a un producto de salida no convencional, como, por ejemplo, que aparezcan ordenados por rango de relevancia o que no sean referencias bibliográficas.

Stephen Harter habla de una visión más restringida de la recuperación de información al considerar solamente los sistemas de recuperación de información en línea accesibles públicamente; así

[...]recuperación de información *online*, o búsqueda *online*, es un proceso en el que un ser humano utiliza un terminal informático para interactuar con un servicio de búsqueda, en un intento de satisfacer una necesidad de información.¹⁴

12 F. W. Lancaster; A. J. Warner, *Information Retrieval Today*. Arlington, Virginia: Information Resources Press, 1993, p. 11.

13 F. W. Lancaster, “Sistemas avanzados de recuperación de información”, en Wilfrid Lancaster y María Pinto (Coords), *Procesamiento de la información científica*. Madrid: Arco/Libros, 2001, p. 213.

14 S. P. Harter, *Online information retrieval: Concepts, principles and techniques*. San Diego: Academic Press, 1986, p. 3.

Geraldene Walker manifiesta que «[...]recuperación de información es el término utilizado comúnmente para describir muchos tipos de búsqueda de literatura».¹⁵

Karen Sparck Jones y Peter Willett indican que la recuperación de información es considerada como sinónimo de recuperación de documentos, y en la actualidad, como recuperación de texto, y que implica dos actividades relacionadas, aunque diferentes: la indización, referida a la representación de los documentos y la petición de información, y búsqueda.¹⁶

Por último, Robert R. Korfhage señala que

[...]almacenamiento y recuperación de la información son dos caras de una misma moneda: si una persona es capaz de buscar información, esa información se ha debido almacenar de alguna manera...Buscar en una colección de documentos puede ser más fácil o más complicado según como se haya organizado la colección.¹⁷

Del análisis de todos estos conceptos y de la lectura de los libros, se pueden extraer las siguientes conclusiones:

1. En ningún libro existe una definición clara y precisa del término *Information Retrieval* (recuperación de la información).
2. No hay un concepto único sobre el significado y lo que abarca la recuperación de información. Para Salton, Van Rijsbergen, Baeza-Yates y Sparck Jones, la recuperación de información incluye tanto la representación como la búsqueda de información. Korfhage distingue entre almacenamiento (representación, organización y almacenamiento de la información) y recuperación (asociada a la búsqueda), y es el único que emplea el término "Almacenamiento de la información y recuperación" (*Information Storage and Retrieval*) en lugar de "Recuperación de la información" (*Information Retrieval*). Lancaster, Walker y Harter relacionan recuperación de información con el proceso de búsqueda, destacando que este último utiliza como sinónimo de recuperación en línea el término "búsqueda en línea". Por último, Meadow también asocia la recuperación con la búsqueda de información, aunque para realizar ésta se ha tenido que organizar y representar previamente la información.

15 G. Walker; J. Janes, *Online Retrieval. A Dialogue of Theory and Practice. Second Edition*, Englewood, Colorado: Libraries Unlimited, Inc., 1999, p. 18.

16 K. Sparck Jones; P. Willett, *Readings in Information Retrieval*, San Francisco: Morgan Kaufman Publishers, Inc., 1997, p. 1.

17 R.R. Korfhage, *Information Storage and Retrieval*. New York: John Wiley, 1997, pp. 1-2.

3. Existe coincidencia total en que el objetivo de la recuperación es proporcionarle información relevante al usuario para satisfacer su necesidad de información.

Observamos, por consiguiente, que existen dos corrientes o tendencias que podemos considerar representativas del pensamiento de los diferentes investigadores y estudiosos del campo de la recuperación de información: (a) aquellos que consideran la recuperación como un proceso de búsqueda, y (b) aquellos que la consideran en un contexto más amplio que engloba tanto la búsqueda como la representación y el almacenamiento de la información.

Estas tendencias también reflejan diferentes audiencias a las que van dirigidos los libros. Así, Walker¹⁸ y Harter¹⁹ declaran en los prefacios de sus libros que éstos se dirigen a estudiantes y profesionales de biblioteconomía y ciencia de la información (*library and information science*), y que su perspectiva es la del usuario y no la del diseñador de sistemas; Lancaster²⁰ manifiesta que el objetivo de su libro es servir de texto en cursos del área de la recuperación de información en general, aunque confía en que sea de utilidad e interés para bibliotecarios, productores de bases de datos y aquellos profesionales interesados en los problemas de acceso temático. Aunque su libro pueda ir dirigido a otras profesiones lo consideraremos dentro de esta tendencia no sólo por el concepto que da sobre recuperación de información, sino también porque su contenido es más acorde con los escritos por los otros autores y por su trayectoria literaria.

La otra tendencia amplía el tipo de público y se centra, aunque sin limitarse a ellos, a departamentos de informática (*computer science*), biblioteconomía (*library science*) y ciencia de la información (*information science*). Así por ejemplo, Meadow²¹ declara que se dirige a estudiantes de ciencias de la información, informática, biblioteconomía, comunicación o periodismo, o a profesionales de servicios de información, y presuponen ellos conocimientos de matemáticas e informática. Baeza-Yates²² declara que su libro se puede utilizar en diferentes áreas como ciencias de la información, informática y biblioteconomía, pero aclara que se centra especialmente en la recuperación de información desde la perspectiva del informático (algoritmos informáticos y técnicas utilizadas en los sistemas de recuperación de información) que difiere de la de bibliotecarios e investigadores de la ciencia de la información

18 G. Walker, *op. cit.*, Prefacio, p. xii.

19 S.P. Harter, *op. cit.*, Prefacio, p. x-xi.

20 F.W. Lancaster; A. J. Warner, *op. cit.*, Prefacio, p. xvii-xviii.

21 C.T. Meadow, *op. cit.*, Prefacio, p. xvi.

22 R. Baeza-Yates, *op. cit.*, Prefacio, p. v.

quienes se centran más en la interpretación humana que se centra en el problema de recuperación de la información.

El libro de Van Rijsbergen²³ va dirigido a estudiantes de ciencia de la información (o informática), biblioteconomía, e investigadores del campo de la recuperación de información, pero hace constar que algunos capítulos requieren conocimientos matemáticos avanzados. Salton²⁴ declara que su texto va dirigido a estudiantes de informática, ciencia de la información y a la gestión. Sparck Jones y Peter Willett²⁵ manifiestan que su libro está dirigido a una audiencia académica, especialmente a departamentos de informática, biblioteconomía y ciencia de la información, y a una audiencia profesional formada por trabajadores de un amplio espectro de organizaciones (bibliotecas, servicios de información, vendedores de bases de datos, compañías multimedia, etcétera) y, sobre todo, a aquellos que se dedican a diseñar e implementar sistemas de recuperación de información. Y Korfhage,²⁶ aunque explícitamente no se dirige a un público concreto, sostiene que el texto se desarrolla principalmente alrededor de métodos informáticos y algorítmicos, esperando que el lector tenga conocimientos fundamentales de conjuntos, lógica, vectores, probabilidad o estadística y geometría.

A la luz de estos comentarios podemos decir que ambas tendencias conceptuales sirven fundamentalmente a tres tipos de audiencia, tanto académica como profesional, representadas en tres ciencias: biblioteconomía (*library science*), ciencia de la información (*information science*) e informática (*computer science*).²⁷ Esto viene a confirmar la naturaleza interdisciplinar del campo de recuperación de información, y a ubicarla esencialmente dentro de tres ciencias más generales.

RELACIÓN DE LA INFORMÁTICA Y LA RECUPERACIÓN DE INFORMACIÓN

La disciplina de la ciencia informática, que nace a principios de la década de 1940 con la confluencia de la teoría de algoritmos, la lógica matemática y la invención de programas almacenados en ordenadores electrónicos, es definida por Peter Denning como

23 C.J. Van Rijsbergen, *Information retrieval*, op. cit., Prefacio a la primera edición, p. vii.

24 G. Salton, op. cit., Prefacio, p. xii.

25 K Sparck Jones; P. Willett, op. cit., Prefacio, p. xi.

26 R.R Korfhage, op. cit., Prefacio, p. xiii.

27 Al respetar la terminología utilizada por los autores se plantea un problema obvio como es la falta del término "documentación", ya que como tal, no se utiliza en Estados Unidos. Considerando que en España el término "ciencia de la documentación" incluye tanto a la biblioteconomía como a la documentación, podríamos decir que *library science* equivale a ciencia de la documentación.

[...]el estudio sistemático de los procesos algorítmicos que describen y transforman información: su teoría, análisis, diseño, eficiencia, implementación y aplicación. La cuestión fundamental subyacente en todo lo que es informática es: ¿Qué puede ser (eficientemente) automatizado?.²⁸

La ciencia informática puede dividirse en varias áreas, e incluir cada una de ellas importantes teorías, experimentos y aspectos relacionados con el diseño, entre estas áreas pueden destacarse las siguientes:

- Algoritmos y estructuras de datos
- Lenguajes de programación
- Arquitectura
- Sistemas operativos y redes
- Ingeniería de software
- Sistemas de recuperación de información y bases de datos
- Inteligencia artificial y robótica
- Gráficos
- Interacción hombre-máquina
- Ciencia computacional
- Informática de organizaciones
- Bioinformática

Es evidente que algunas de estas áreas son propias de o están relacionadas con la recuperación de información; naturalmente, el enfoque y los problemas que trata de resolver la informática están orientados al diseño de aplicaciones y de sistemas informáticos tanto de software como de hardware.

El área etiquetada como *Sistemas de recuperación de información y bases de datos* se ocupa fundamentalmente de la organización de los datos en el sistema, los modelos de representación, las operaciones que se realizan para el almacenamiento, la localización, la equiparación y recuperación, la interacción del usuario, el diseño de arquitecturas para una recuperación más eficiente, etcétera. Así, el área trata de las técnicas para diseñar SRI (Sistemas de recuperación de información) y bases de datos.

El área de *Inteligencia artificial y robótica* se propone modelar el conocimiento humano con la intención de construir componentes o sistemas automatizados que imiten el comportamiento humano o aumenten sus capacidades. Conceptos como proceso en lenguaje natural, redes semánticas, redes

28 P. J. Denning, "Computer Science: The discipline", en A. Ralston y D. Hemmendinger (Eds), *Encyclopedia of Computer Science*, 2000 Edition. London: Nature Publishing Group, 2000, p. 1. <<http://www.idi.ntnu.no/~ekaterip/dif8916/denning.pdf>>

neuronales, modelos de aprendizaje, agentes inteligentes y robots en la web, pertenecen al área de investigación en inteligencia artificial relacionada con la recuperación de información.

Otra área relacionada directamente con la recuperación de información, y de especial interés en los últimos años, es la *Interacción hombre-máquina* (HCI), que se ocupa fundamentalmente del diseño de interfaces y en la que también intervienen el área de *Gráficos informáticos* y otras como la *Psicología cognitiva*.

Otras áreas que han desempeñado un papel fundamental en la recuperación de información han sido *Algoritmos y estructuras de datos*, *Lenguajes de programación*, *Arquitectura y sistemas operativos*, y *Redes*. Los dos últimos han contribuido a que los ordenadores sean cada vez más potentes, rápidos, baratos y a que se hayan convertido en una herramienta personal de trabajo utilizada en muchas ocasiones para la búsqueda y recuperación de información, la que facilita el proceso de comunicación y difusión de la información. Con respecto a la primera área, la historia de la recuperación de información está plagada de la implementación de diversos algoritmos para la entrada, comparación y salida de los resultados.

La mayor parte de la literatura e investigación en informática se ocupa del ordenador digital y de los fenómenos que lo rodean, como la estructura y el funcionamiento de los sistemas informáticos, los principios de diseño y la programación de ordenadores, los métodos eficaces para usar los ordenadores en tareas de proceso de información, y las teorías sobre sus propiedades y limitaciones. El campo de estudio son los propios ordenadores, a diferencia de otros campos, donde éstos son simplemente herramientas.

Esta última frase perfila claramente la relación entre la ciencia de la información y de la documentación con la informática, y se basa en la aplicación de los ordenadores, junto con sus productos y servicios asociados, para recuperar información. Pero hay otra diferencia que conviene destacar y que señala Saracevic: la ciencia informática trata de algoritmos relacionados con la información, considerada ésta como señales o mensajes (manipulación de símbolos), mientras que la ciencia de la información y documentación trata sobre la propia naturaleza de la información y su uso por los seres humanos (manipulación del contenido) en un contexto.²⁹

Desde un principio, la recuperación de información ha estado ligada a la ciencia de la informática no sólo por el uso de ordenadores y de las tecnologías de la información como herramientas de trabajo, sino también porque una parte importante de la investigación ha estado orientada al diseño de mejores

29 T. Saracevic, "Information science", en *Journal of the American Society for Information Science*, 1999, vol. 50, nº 12, p. 1060.

sistemas de recuperación de información. Para Baeza-Yates, el problema de la recuperación de información desde el punto de vista de la informática consiste principalmente en diseñar y construir índices eficientes, en procesar las consultas de los usuarios con un alto rendimiento, y en desarrollar algoritmos de rango que mejoren la calidad de los resultados recuperados.³⁰

Estas soluciones para tratar el problema de la recuperación de información también han sido estudiadas por la ciencia de la información en lo que se ha denominado como enfoque centrado en los sistemas³¹ o la tradición informática.³² Así, pioneros como Calvin Mooers, Mortimer Taube, James Perry, Allent Kent, Hans Peter Lung, Maron, Kuhns, Gerard Salton, etcétera, estaban interesados en el diseño de nuevos y mejores índices y sistemas de recuperación de información, y en el desarrollo de algoritmos matemáticos y probabilísticos. Muchos de ellos eran ingenieros, informáticos o matemáticos que se convirtieron en personalidades no en el campo de la informática sino en el de la ciencia de la información.

En la actualidad todavía persisten los problemas de software y hardware asociados con la recuperación y la representación de documentos. El desarrollo de SRI digitales requiere de programas informáticos que comparen las consultas con los documentos almacenados en las bases de datos, y luego produzcan una salida. En estos sistemas, tanto los términos de entrada como el texto de salida se pueden ordenar según ciertos criterios; el reto para los investigadores del enfoque de sistemas es desarrollar algoritmos que optimicen estos rangos.³³

Otras líneas de investigación propias de la informática son los sistemas de información para la gestión, los sistemas de gestión de bases de datos, la inteligencia artificial y los sistemas expertos, el hipertexto, la interacción hombre-máquina y, más recientemente, las bibliotecas digitales. Todo esto se concentra en el Special Interest Group on Information Retrieval (SIGIR) de la Association for Computing Machinery (ACM) y en algunos otros como el Special Interest Group on Computer-Human Interaction (SIGCHI),³⁴ sin embargo, es curioso que todos estos temas sean abordados también por los Special Interest Groups de la American Society for Information Science & Technology

30 R. Baeza-Yates; B. Ribeiro-Neto, *op. cit.*, p. 7.

31 T. Saracevic, *op. cit.*, pp. 1051-1063.

32 M. Buckland, "The landscape of information science: The American Society for Information Science at 62", en *Journal of the American Society for Information Science*, 1999, vol. 50, n° 11, pp. 970-974.

33 D. Robins, "Interactive information retrieval: Context and basic notions", en *Informing Science*, 2000, vol. 3, n° 2, p. 60.

34 Véase el sitio web de ACM SIGIR (<http://www.acm.org/sigir/>) donde se tratan todos los temas relacionados con la aplicación de los ordenadores en la adquisición, organización, almacenamiento, recuperación y difusión de la información.

(ASIST), y así observamos que hay un grupo que trabaja sobre Digital Libraries (DL), y otros sobre Computerized Retrieval Services (CRS), Human-Computer Interaction (HCI) o Knowledge Management (KM), y algunos más otros.³⁵

Existen por lo tanto dos organizaciones diferentes pero que investigan temas comunes a la recuperación de información, ¿cuál es la diferencia entre ellas? ¿Podría decirse por el nombre, que la ACM representa la parte informática pura y dura relacionada con la información, mientras que ASIST representa la parte de ciencia de la información relacionada con la informática? Saracevic habla de la división de enfoques, uno centrado en los sistemas y representado por SIGIR de la ACM, y el otro centrado en los usuarios y congregado alrededor de ASIST.³⁶ Para este autor ambos enfoques tratan a la recuperación de información desde dos perspectivas diferentes; el enfoque de sistemas está representado por el trabajo sobre algoritmos y evaluación basada en el modelo tradicional de recuperación de información; y el enfoque centrado en los usuarios se interesa en el uso de los sistemas de recuperación de información.

Ahora bien, si uno lee los objetivos de ambos grupos no parece que esté tan clara esta diferencia. Es más, cuando Ellis habla del paradigma centrado en los sistemas y del paradigma centrado en los usuarios, se refiere únicamente a los diferentes enfoques que existen dentro de la ciencia de la información al tratar los problemas de la recuperación de información.³⁷ Por lo tanto, parece que existe un área de investigación de la recuperación de información que comparte intereses comunes tanto con la ciencia de la información como con la ciencia de la informática, pero los límites con ambas son muy confusos y pertenecen a dos organizaciones diferentes.

El desafío está en integrar y hacer que trabajen juntos ambos enfoques; no se pueden diseñar nuevos y mejores sistemas de recuperación de información si no se tienen en cuenta las características y el comportamiento de los usuarios a quienes están dirigidos dichos sistemas. Son muchas las líneas de investigación actuales que pueden atraer a ambos mundos y establecer una colaboración multi e interdisciplinar, tales como la interacción hombre-máquina con el diseño de interfaces y la creación de SRI inteligentes, en las cuales es necesario entender cómo utilizan los usuarios las estrategias y términos de búsqueda, cuáles resultados son juzgados como relevantes por los usuarios para el mismo problema de búsqueda, etcétera. Esta nueva área de investigación, conocida como recuperación de información interactiva, requiere una

35 Véase <http://www.asis.org/AboutASIS/asis-sigs.html#SIGHFIS>, donde se especifican todos los grupos y sus objetivos.

36 T. Saracevic, "Information science...", *op. cit.*, p. 1057.

37 D. Ellis, *Progress and problems in Information Retrieval*. London: Library Association Publishing, 1996.

colaboración interdisciplinar, porque el problema es demasiado grande para ser resuelto por un sólo campo y requiere de investigadores de diferentes disciplinas como psicología, sociología, comunicaciones, ciencia de la información o informática, por citar unas pocas.³⁸

RELACIÓN DE LA CIENCIA DE LA INFORMACIÓN CON LA RECUPERACIÓN DE INFORMACIÓN

La relación e influencia entre la ciencia de la información y la recuperación de información es de tal magnitud que podemos decir que esta última es uno de los antecedentes directos de la ciencia de la información. Así lo manifiesta Saracevic cuando dice que «[...]la Recuperación de Información no fue la única responsable del desarrollo de la Ciencia de la Información, pero sí la más importante».³⁹ Hay que recordar que la recuperación de información empieza a cobrar identidad a raíz del famoso artículo publicado en 1945 por Vannevar Bush⁴⁰ sobre los problemas generados por el crecimiento y la explosión de la información científica, y al ser acuñado el término por Calvin Mooers en 1950. El comienzo de las investigaciones para diseñar nuevos mecanismos y sistemas de recuperación de información automatizados que resolvieran el problema de la información marca un periodo de tiempo en el que la recuperación de información centra el interés de científicos, investigadores y profesionales de distintas disciplinas y adquiere una gran importancia. Tanto es así, que en el análisis que realiza Schrader de los términos utilizados en Estados Unidos para denominar a la disciplina, ofrece los siguientes.⁴¹

1900 a 1930	Bibliografía
1920 a 1950	Documentación
1940 a 1950	Información científica
1950 a 1960	Recuperación de la información
1960-	Ciencia de la información

y destaca que en la década de los 50 se utilizaba el término recuperación de la información, pero que posteriormente pasó a denominarse ciencia de la información.

38 D. Robins, *op. cit.*, p. 60.

39 T. Saracevic, "Information science: origin, evolution and relations", *op. cit.*, p. 9.

40 V. Bush, "As we may think", en *Atlantic Monthly*, 1945, 176/1, pp. 101-108.
<<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush0.shtml>>

41 A. M. Schrader, "In search of a name: information science and its conceptual antecedents", en *Library and Information Science Research*, 1984, vol. 6, n° 3, pp. 227-271.

Resulta evidente, por tanto, que el nacimiento de la ciencia de la información supone la fagocitación y absorción de la recuperación de información, y hace suyos sus problemas, métodos y técnicas, erigiéndose desde sus inicios en su principal área de investigación.

Moreiro destaca que el primer texto que estudió la ciencia de la información como disciplina independiente se titulaba *Almacenamiento y recuperación de la información*, que además de reflejar la variedad que caracterizó a esta ciencia desde un principio, destacaba sobre todo que la recuperación de información se había convertido en la auténtica protagonista de la actividad documental.⁴²

También hay un aspecto que conviene destacar, y es que para Schrader el término documentación se utiliza de 1920 a 1950, y ya hemos comentado anteriormente que en Estados Unidos el término desaparece para dar lugar a Information Science, pero hay que recordar que el movimiento documental en ese país empieza a tener peso específico por el uso de los ordenadores y los sistemas de recuperación de información, sobre todo en las bibliotecas y centros de documentación, y que por tanto existe un vínculo estrecho entre documentación y recuperación de información.

Por lo tanto, la documentación (en Estados Unidos) es también absorbida por la ciencia de la información, si bien no en todos sus procesos, sí en aquellos relacionados con la recuperación de información, y es en este sentido que se expresa López Yepes cuando dice:

Borko es el primer autor que fija el concepto de Ciencia de la Documentación en el campo de la Information Science, especialmente en la fase de la recuperación de información.⁴³

De igual manera se expresa Kochen, para quien la ciencia de la información es el alma del trabajo científico, se deriva en su origen de la documentación y de la recuperación de información.⁴⁴

La ciencia de la información es, por lo tanto, un campo relacionado con los problemas que surgen en la comunicación a partir de la información y del conocimiento en general, y con los registros de esta comunicación en particular, pero comparte estos intereses con otros campos, especialmente la biblioteconomía y la documentación, lo cual establece una estrecha relación

42 J. A. Moreiro, *Introducción al estudio de la información y la documentación*. Medellín: Editorial Universidad de Antioquía, 1998, p. 101.

43 J. López Yepes, *La documentación como disciplina...*, *op. cit.*, p. 166.

44 M. Kochen, "Views on the foundations of information science", en: Anthony Debons (Ed). *Information Science: Search for identity*. New York: Marcel Dekker, Inc., 1974, pp. 171-177.

entre ellos.⁴⁵ Esta relación ha sido de gran interés para la comunidad académica y profesional, pero existe todavía cierta incertidumbre sobre su relación. Así, hay una concepción que considera a la biblioteconomía y a la documentación como disciplinas separadas de la ciencia de la información, pero con intereses comunes. La otra concepción es la que las considera idénticas y formando un único campo. Una variante de esta última visión es considerar que un campo es parte del otro.⁴⁶

La ciencia de la información, como cualquier otro campo, está estructurada en diferentes áreas o subdisciplinas de investigación o práctica. Para examinar y analizar esta estructura, nos basaremos en el trabajo realizado por White y McCain en el que realizan un análisis de co-citas sobre los 120 autores más citados en el periodo comprendido entre 1972 y 1995, en 12 revistas de ciencia de la información y automatización de bibliotecas.⁴⁷

Los autores identifican dos grandes grupos o *clusters* de autores que representan dos áreas principales:

1. Primer grupo, denominado *domain cluster* o área básica, representado por estudios que se centran en el fenómeno de la información y sus manifestaciones en la literatura. Está formado por autores que trabajan en el estudio analítico de literaturas; sus estructuras; estudios de textos como portadores de contenido; comunicación, especialmente la comunicación científica; contexto social de la información; usos de la información; búsqueda de información y comportamiento; diferentes teorías de la información y temas relacionados.
2. Segundo grupo, *cluster* de recuperación o área aplicada, que aborda fundamentalmente la implementación, el comportamiento y los efectos de las interfaces entre documentos y personas, incluyendo todos los diferentes aspectos de la recuperación. Los autores trabajan en teorías y algoritmos de recuperación de información; procesos y sistemas de recuperación de información en la práctica; interacción hombre-máquina; estudios de usuarios; sistemas bibliotecarios; OPACS y temas relacionados.

45 T. Saracevic, "Relevance: A review of and a framework for the thinking on the notion in information science", en: Karen Sparck Jones and Peter Willett (Eds). *Readings in Information Retrieval*, San Francisco, California: Morgan Kaufman, 1997, p. 145.

46 P. Vakkari, "Library and Information Science: Content and scope", en Johan Olaisen, Erland Munch-Petersen, and Patrick Wilson (Eds), *Information Science. From the development of the discipline to social interaction*, Oslo: Scandinavian University Press, 1996, pp. 169-231.

47 H.D. White; K. W. McCain, "Visualizing a discipline: An author co-citation analysis of Information Science, 1972-1995", en *Journal of the American Society for Information Science*, 1998, vol. 49, n° 4, pp. 327-355.

De este análisis de *clusters* se destaca la falta de conexión entre estos dos grupos y que el grupo de recuperación es mucho más numeroso, tanto en número de autores como en número de trabajos.

Además de este análisis estructural en dos grandes grupos, se desglosan la ciencia de la información en áreas específicas mediante un análisis factorial de autores co-citados, del que se extraen doce factores identificados como especialidades y a los que se les asignan las siguientes etiquetas:

- Recuperación de información experimental
- Análisis de citas
- Recuperación en línea (práctica)
- Bibliometría
- Teoría de sistemas bibliotecarios, incluida la automatización de bibliotecas
- Comunicación científica
- Teoría y estudios de usuarios
- OPACS (On-line public Access Catalogs)
- Ideas importadas de otras disciplinas, como teoría de la información, informática, ciencia cognitiva
- Teoría de la indización
- Teoría de citas
- Teoría de la comunicación

Estas especialidades dan una visión bastante exacta de aquellas áreas en las que han trabajado los científicos de la información hasta 1995. Evidentemente, desde esta fecha hasta la actualidad han surgido nuevas áreas, (incluso antes, ya que el trabajo de estos autores se basa en autores co-citados y es de sobra conocido que desde que se publica un trabajo hasta que se cita puede pasar un tiempo razonablemente largo), y así Saracevic sugiere que las principales áreas nuevas incluyen estudios de interacción, búsqueda en Internet, recuperación de información multimedia, recuperación de información multilinguaje y bibliotecas digitales.⁴⁸

De la división de la ciencia de la información en las especialidades enumeradas anteriormente, conviene destacar varios aspectos:

1. La existencia de dos especialidades que contienen el término recuperación de información, una experimental centrada en el diseño y evaluación de sistemas de recuperación de información, y otra práctica o

en línea que se centra fundamentalmente en la recuperación en bases de datos reales.

Estas dos especialidades concuerdan con los dos tipos de conceptos de recuperación de información que hemos proporcionado anteriormente en el apartado de análisis de conceptos, donde la recuperación de información experimental estaría relacionada con el concepto amplio que incluye tanto la entrada y la salida en el sistema como el propio SRI, mientras que la recuperación práctica sólo se centraría en las técnicas de búsqueda y el uso de bases de datos reales. Así pues, esta distinción resultará esencial para delimitar la recuperación de información en el ámbito de la ciencia de la documentación, donde la formación de un futuro experto en este campo deberá tener un marcado componente práctico en lugar del componente científico o experimental.

También coincide la clasificación de los autores de esos conceptos con las especialidades mencionadas; así, Salton y Van Rijsbergen pertenecen a la recuperación experimental, mientras que Meadow aparece en la recuperación práctica. Lancaster y Harter aparecen en ambos factores, lo que para White y McCain significa que tienen contribuciones en ambas especialidades y que han podido emigrar de una a otra especialidad. Así por ejemplo, Lancaster, por su famosa evaluación del sistema MEDLARS, era considerado en un principio como un recuperador experimental y posteriormente como un recuperador práctico.

2. La existencia de estas dos especialidades en recuperación de la información no significa que las demás no estén relacionadas con ella. Una de las más importantes, y que está implicada directamente en la recuperación de información, es la especialidad etiquetada como *Teoría y necesidades de usuarios*, encabezada por Dervin. Varios autores clasificados en esta especialidad también se incluyen en la de recuperación experimental (Belkin, Swanson, Oddy, Saracevic, Ingwersen, P. Wilson) y en la de recuperación práctica (Borgman, Fidel, Bates), y sólo Dervin, T.D. Wilson y R. Taylor puntúan exclusivamente en este *cluster*. Esta especialidad corresponde al nuevo enfoque que se da en la década de los 80 en el campo de la recuperación de información, el enfoque cognitivo, centrado en las necesidades de información de los usuarios y en su interacción con el sistema; así, en los trabajos de estos autores es frecuente que aparezcan términos como estrategias de búsqueda, formulación de preguntas, estilos de búsqueda de información y juicios de relevancia.

Otra especialidad implicada directamente con la recuperación de información es la denominada *OPACS*; de hecho, los *OPACS* son un sistema de

recuperación de información de las bibliotecas. Así, autores que tienen que ver con esta especialidad (Cochrane, Matthews, Hildreth) también lo hacen en la recuperación práctica, y autores que se relacionan con la recuperación práctica (Borgman y Fidel) también lo hacen en OPACs.

Del estudio de White y McCain se extrae una conclusión, que resume muy bien Saracevic:

[...] la recuperación de información ha sido, y sigue siendo, la principal área de la ciencia de la información y en la que mayor número de autores y trabajos están implicados...Seguramente, la ciencia de la información es más que recuperación de información, pero muchos de los problemas tratados por la RI o derivados de objetos y fenómenos implicados en RI, están en su núcleo.⁴⁹

RELACIÓN DE LA CIENCIA DE LA DOCUMENTACIÓN Y RECUPERACIÓN DE INFORMACIÓN

Para ubicar la recuperación de información dentro de la ciencia de la documentación y examinar sus relaciones, debemos abordar en primer lugar el concepto de documentación. José López Yepes, en la última edición de una obra ya clásica, nos presenta dos definiciones del término documentación.⁵⁰

A) Como conjunto de las disciplinas documentarias que estudian y ejecutan los diversos aspectos del proceso documental (Archivística, Biblioteconomía, Documentación, Museología, Bibliografía, etcétera). En este sentido, la denominación equivaldría a la de Ciencias de la Documentación.

En este sentido,

[...] representa el conjunto de disciplinas que tienen por objeto el estudio de un proceso informativo en el que se da una actividad de recuperación de mensajes emitidos en procesos anteriores y que, mediante análisis y tratamiento técnico, se comunican transformados con la finalidad de que sirvan de fuente de información para la obtención de nuevo conocimiento o para la toma de decisiones.

B) Como disciplina responsable del estudio de una parte del proceso documental.

49 *Ibid.*, pp. 1056-1057.

50 J. López Yepes, *op. cit.*, pp. 321-322.

En este sentido

Documentación es una de las Ciencias de la Documentación que tiene como objeto de estudio aquella parte del proceso documental consistente en la recuperación y difusión de mensajes documentarios y su aprovechamiento por parte del sujeto receptor o usuario a fin de que sirvan de base para la obtención de nuevo conocimiento o para la toma de decisiones.

En ambas definiciones aparece el término “recuperación”. En la primera, como actividad común que se realiza en todas las disciplinas de las ciencias de la documentación, pero aún podemos ir más allá siguiendo a Lancaster, quien sostiene que es la principal actividad en los centros o servicios de información, incluidas las bibliotecas o cualquier otro tipo de servicio que proporcione recursos de información a una población de usuarios.⁵¹ La segunda definición coincide casi plenamente con los conceptos y el objetivo de la recuperación de información que se han expuesto anteriormente.

Otros autores que han estudiado en profundidad el concepto de documentación son Sagredo e Izquierdo, quienes definen, de una manera abierta, a la documentación como

[...] el proceso de comunicación por el que un individuo (persona o colectividad) recibe las referencias de unos objetos (los documentos) debidamente tratados – opcionalmente también esos mismos objetos referidos– expresa o presuntamente pedidos (demandados), relativos a un determinado dominio de la actividad social que el individuo desempeña, tales que le hacen competente para el ejercicio de esa actividad, cuando atinadamente los maneja.⁵²

Esta definición de documentación es claramente coincidente con el amplio concepto de recuperación de información, similaridad que todavía es más marcada si comparamos los componentes que según estos autores definen a la documentación con los elementos que intervienen en el proceso de recuperación de información,⁵³ como se muestra en la siguiente tabla:

51 F. W. Lancaster; A. J. Warner., *op. cit.*, p. 1.

52 F. Sagredo Fernández; J. Ma. Izquierdo Arroyo, *Concepción lógico-lingüística de la documentación*. Madrid: IBERCOM-Red Comnet de la Unesco, 1983, pp. 286-287.

53 J.M.Tagüe-Sutcliffe, *op. cit.*, pp. 467-490.

Componentes definicionales de documentación según Sagredo e Izquierdo	Componentes del proceso de recuperación de información
Agente individual ("documentador")	Intermediario (documentalista, bibliotecario o profesional de la información)
Receptor individual (persona o colectividad)	Usuario o comunidad de usuarios
Un mensaje (texto lingüístico oral o de otro modo canalizado) que refiere (habla de) y está localizado espacialmente (en un lugar: biblioteca, archivo,...)	Documentos o información contenida en bases de datos
Objeto o conjunto de objetos debidamente tratados (documentos)	Documentos tratados y representados en la base de datos (indización, resumen, descripción, etc.)
Una intención comunicativa específica por parte del agente: hacer llegar al receptor el mensaje que pone a disposición de éste (como corpóreamente dados ya, o simplemente como referidos) ese objeto o conjunto de objetos	El intermediario proporciona al usuario las referencias o el texto completo de los documentos recuperados en la búsqueda
Una presuposición, por parte del agente, de que éstos son los objetos que le interesan al receptor (por haberlos así expresamente señalado, o por conclusiones deducidas del estudio de su "perfil" socio-ocupacional)	A través de la entrevista de referencia, el intermediario debe conocer la verdadera necesidad de información del usuario para diseñar la estrategia de búsqueda adecuada que le permita recuperar aquellos documentos pertinentes, y poder modificar la estrategia de búsqueda hasta conseguir los resultados esperados
Un proceso (enmarcado en los componentes anteriores) comunicativo en esencia, que presupone acciones no comunicativas (tales como búsqueda, denominación, cotejo, etc.)	El proceso de recuperación de información es siempre un proceso de comunicación ⁵⁴
Una correspondencia específica entre la presuposición y la intención receptora	La estrategia de búsqueda debe de corresponder con la necesidad de información real del usuario, y es, sin duda, uno de los factores que influyen en el éxito de la recuperación de información
La inserción del receptor en una determinada esfera social (coincidente o no con la del documentador).	<i>Idem.</i>
Un cambio de estado por parte del receptor, progresivo y a su favor, consistente en pasar de la carencia a su posesión.	<i>Idem.</i>
El uso o manejo apropiado, por parte del receptor, de esas referencias y/u objetos	<i>Idem.</i>

En el análisis que realizan Sagredo e Izquierdo sobre el concepto científico de la documentación a partir de cinco perspectivas otletianas, definen a la Documentación como «[...]ciencia que estudia la naturaleza del documento y su proceso de tratamiento y recuperación».⁵⁵

54 A diferencia de lo que expresan Sagredo e Izquierdo, cuando dicen que el proceso de la documentación no es puramente comunicativo, el proceso de recuperación de información implica siempre un proceso de comunicación, ya que la búsqueda implica comunicación entre el intermediario o buscador con el SRI, o incluso el cotejo o equiparación, pues aunque éste es un proceso mecánico, puede implicar indirectamente la interacción (diálogo interactivo) del buscador con el sistema para modificar la estrategia de búsqueda (*feedback* por relevancia manual o automático) y conseguir mayor precisión o llamada.

55 F. Sagredo Fernández; J.M^o. Izquierdo Arroyo, *op. cit.*, p. 418.

De lo anteriormente expuesto, resulta evidente que el término y concepto de “recuperación” aparece de forma explícita o implícita en los conceptos de documentación, y que contribuye en gran manera a que la documentación sea «[...]una especie de ciencia para la ciencia, de ciencia instrumental al servicio de la creación y propagación de las conclusiones científicas...»,⁵⁶ ya que el objetivo último de la recuperación de información es satisfacer las necesidades de información de los usuarios para que éstos puedan servirse de ella y crear nuevo conocimiento, objetivo que coincide con el de documentación como ciencia, y así se expresa Moreiro al afirmar que ésta «[...] realiza un proceso de transformación tendente a ofrecer a los usuarios la información que precisen mediante una recuperación eficaz y exacta».⁵⁷

El desarrollo histórico y teórico de la documentación ha sido objeto de numerosos estudios y no pretendemos aportar nuevas contribuciones (algo que, además, sería hartamente difícil) sino conocer las manifestaciones de la recuperación de información y su influencia en la documentación.

Paul Otlet, traité de documentation y recuperación de la información

Los orígenes de la documentación se remontan a finales del siglo XIX gracias a la acción de dos abogados belgas, Paul Otlet y Henri La Fontaine, considerado el primero de ellos como “el padre de la documentación”, a raíz de cuyas investigaciones comienza la institucionalización del movimiento documental, ya que la documentación, fuente primordial de información, comienza a ser estudiada desde un punto de vista técnico, es decir, el tratamiento lógico que se ha de seguir en la búsqueda y recuperación de la masa documental.⁵⁸ En el año 1895 se creó el “Instituto Internacional de Bibliografía”, antecedente del actual “Instituto Internacional de Documentación”, con la idea de resolver el problema de la información de las fuentes en la investigación científica que se debían al crecimiento de la literatura; estos avances desbordaron, en opinión de M^a Luz Terrada, los cauces de la bibliografía y condujeron a construir la documentación como disciplina en respuesta a las necesidades informativas características de la actividad científica contemporánea.⁵⁹

56 J. López Yepes, “Introducción”, en: José López Yepes (Coordinador). *Manual de Información y Documentación*. Madrid: Pirámide, 1996, p. 23.

57 J. A. Moreiro González., *Introducción bibliográfica y conceptual al estudio evolutivo de la Documentación*. Barcelona: PPU, 1990, pp. 308-309.

58 M. Caridad Sebastian, *Fundamentos Teóricos en Documentación Automatizada: Programa y Bibliografía*. Barcelona: PPU, 1993, p. 11.

59 M^a. L.Terrada; J. M^a. López Piñero, “Historia del concepto de Documentación”, en *Documentación de las Ciencias de la Información*, 1980, vol. 4, p. 238.

Otlet comienza a formular algunas ideas interesantes y en 1934 publica su *Traité de la Documentation*, acontecimiento éste de gran importancia y considerado por Sagredo e Izquierdo como

[...] el punto de partida de nuestro acceso a la concepción científica de la Documentación, entendiéndolo que el eminente promotor belga es el representante de una primera fase en la evolución histórica de nuestra disciplina...⁶⁰

En las páginas finales de esta obra, Otlet propuso tres hipótesis sobre el acceso eficaz a toda la documentación, siendo de especial interés para nosotros la Tercera Hipótesis.

Esta hipótesis presenta la visión futurista de Otlet sobre el acceso automático a distancia a la documentación existente en el mundo con base en los siguientes elementos: una pantalla, un teléfono, un inmenso edificio donde se encuentran todos los libros e información, y personal altamente cualificado para seleccionar y combinar la información. Evidentemente, lo que Otlet profetizó era lo que hoy denominamos teledocumentación o recuperación de información en línea, en la que desde un ordenador y a través de redes de telecomunicaciones, un profesional, o bien el propio usuario, accede a la información contenida en bases y bancos de datos (tratados a su vez por personal especializado) y que aparecen al instante en la pantalla del ordenador.

Para Rayward, el *Traité de Documentation* de Otlet es, en realidad, un tratado sobre almacenamiento y recuperación de información situado en un amplio contexto de prácticas de comunicación y el conocimiento de aquel entonces, donde el término “información” sustituye los fenómenos abstractos y los procesos técnicos y profesionales que Otlet trata.⁶¹ Este autor analiza las ideas y prácticas expuestas en el tratado, y concluye que en la actualidad equivaldrían a tecnología de la información, recuperación de información, estrategias de búsqueda, centros de información, servicios de información de pago, bases de datos vinculadas, software de gestión de bases de datos, redes académicas, multimedia e hipertexto; conceptos todos ellos claramente relacionados con la recuperación de información.

Hasta ahora hemos visto la relación de la recuperación de la información con tres ciencias: documentación, ciencia de la información e informática. La informática se ocupa esencialmente del diseño de los SRI, tanto de hardware como de programas. En la ciencia de la información, había dos subgrupos de

60 F. Sagredo Fernández; J.M^a. Izquierdo Arroyo, *op. cit.*, p.294.

61 W. B. Rayward, “The origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID)”, en *Journal of the American Society for Information Science*, 1997, vol. 48, n^o 4, pp. 289-300.

recuperación de información: 1) recuperación de información en línea, centrada en la búsqueda de información en bases de datos reales, y 2) recuperación de información experimental, que se ocupa del diseño y evaluación de sistemas; en este gran grupo, cabrían dos enfoques o subgrupos, el orientado a sistemas y afín a la ciencia de la informática, y el orientado al usuario, relacionado con ciencias sociales y cognitivas, y en el que se estudia el comportamiento del usuario para diseñar sistemas más eficientes.

Queda por analizar, y es lo que nos proponemos a continuación, el papel que desempeña la recuperación de información dentro de la ciencia de la documentación. Ello nos permitirá ir estableciendo sus límites y avanzando en su definición.

La recuperación de información en el proceso documental

Según el concepto de ciencia de la documentación revisado anteriormente, es necesario contemplar la recuperación de información como actividad o proceso que se produce dentro de uno más amplio conocido como proceso documental. Las fases principales de este proceso documental pueden resumirse en las siguientes: 1) Selección y adquisición, 2) Análisis, 3) Almacenamiento, 4) Búsqueda o recuperación,⁶² y 5) Difusión de la información. Las tres primeras corresponden a la entrada en el sistema, mientras que las dos últimas hacen referencia a la salida.

La importancia de la recuperación de información dentro del proceso documental ha sido resaltada por Martínez Comeche, para quien

[...]la labor de recuperación se erige en una de las tareas más intrínsecamente documentales, pues en ella radica, en gran medida, la utilidad del proceso documental desde el punto de vista del usuario, y porque en ella se consume una de las más altas misiones de la Documentación, la de informar al receptor sobre dónde se halla la información que necesita para generar nuevo conocimiento, contribuyendo de este modo al avance científico en particular y al progreso en general.⁶³

Sin embargo, el propio Martínez Comeche comenta que, a pesar de su trascendencia, hay discrepancias entre los diferentes autores sobre cuál es su

62 El uso indistinto por parte de los autores del término recuperación o búsqueda lo pone de manifiesto Martínez Comeche cuando dice: «[...]la puesta a disposición de los mensajes requiere a veces una actividad previa por parte del documentalista, la denominada recuperación —conocida también por búsqueda o selección— [...]». Véase Martínez Comeche, J.A. “El proceso informativo-documental”, en: López Yepes, J. (Coordinador), *Manual de Información y Documentación*. Madrid: Pirámide, 1996, p. 42.

63 *Ibid.*, p. 44.

posición dentro del proceso informativo documental, debido sobre todo a la vinculación existente entre las tareas de análisis y recuperación. Estas desavenencias han sido bien estudiadas por Pinto Molina⁶⁴ y son consecuencia del uso del término *tratamiento documental*. Algunos autores hablan de tratamiento documental para referirse al conjunto de operaciones de análisis y recuperación, mientras que otros se refieren sólo al análisis, y existe otro grupo que le da un enfoque diferente al asociarlo a la explotación de los fondos documentales.

Coincidimos plenamente con Pinto Molina cuando dice «[...] la expresión “tratamiento” significa básicamente transformación de la información contenida en los documentos a fin de posibilitar su control y recuperación».⁶⁵

Esta transformación implica una serie de actividades encaminadas a representar la información en un SRI, tales como indización, resumen, catalogación o descripción, y que corresponden todas a un análisis físico o formal y a un análisis de contenido, los dos polos que concretan el campo de actuación del análisis documental, disciplina científica dentro de la documentación.

Por lo tanto consideramos que el análisis y la recuperación son dos eslabones del mismo proceso, íntimamente relacionados pero independientes, ya que el objetivo del análisis documental es la transformación y representación de la información para una mejor recuperación.

El análisis formal y de contenido de los documentos son actividades intelectuales realizadas tradicionalmente por bibliotecarios y documentalistas, y esto se refleja en los planes de estudio de biblioteconomía y documentación en España a través de diferentes disciplinas académicas como indización, resumen, catalogación o bibliografía. Sin embargo estas actividades de representación temática como el resumen y, sobre todo, la indización, se pueden realizar también de manera automática mediante la aplicación de diferentes algoritmos. Y ésta ha sido, como hemos señalado con anterioridad, una área de gran profusión investigativa en el campo de la recuperación de información.

Todas estas actividades de análisis de la información, sean realizadas por personas o por algoritmos informáticos, tienen en común que su objetivo es la representación temática de los documentos y que se realizan antes o durante la fase de entrada en el SRI. Por lo tanto, con base en todo lo expuesto hasta aquí, podemos extraer una serie de conclusiones clave para el desarrollo de nuestro concepto:

1. Que la recuperación de la información dentro del proceso documental informativo, y en el marco de la ciencia de la documentación, es

64 M. Pinto Molina, *Análisis documental. Fundamentos y procedimientos* (2ª edición). Madrid: Eudema, 1993, pp. 40-45.

65 *Ibid.*, p. 44.

un proceso que tiene lugar en la salida del sistema y que se asocia a la búsqueda de información.

2. Que todas las actividades de representación, análisis e indización que se realizan durante la entrada en el sistema están relacionadas con la recuperación de información, pero que no forman parte del propio proceso de recuperación o búsqueda.
3. Que estas actividades de entrada son independientes de la recuperación de información, y que en el marco de la ciencia de la documentación corresponden a una disciplina propia denominada análisis documental, y dentro de ésta a subdisciplinas como catalogación, indización y resumen.

Por lo tanto, en el proceso documental y dentro del marco de la ciencia de la documentación delimitamos la recuperación de información de todo aquello que implica análisis, representación y almacenamiento de la información, lo cual está asociado a la búsqueda de información.

DEFINICIÓN DE RECUPERACIÓN DE INFORMACIÓN EN LA CIENCIA DE LA DOCUMENTACIÓN

Llega el momento de presentar una definición de la recuperación de información que, aunque pueda constituir un freno como cualquier otra definición, pretendemos que sea el resultado de todas las reflexiones e ideas que se han presentado en las páginas anteriores; a modo de síntesis podemos destacar las siguientes:

1. La *naturaleza interdisciplinar* de la recuperación de información. Se ha considerado como un campo amplio relacionado con los procesos de representación, organización, almacenamiento, búsqueda y acceso a la información, en el que han intervenido a lo largo de su historia investigadores y estudiosos de diferentes ciencias: como la informática, ciencia de la información, documentación, lingüística, ciencias cognitivas, inteligencia artificial, lógica, etcétera, y aunque desde una perspectiva diferente, todas ellas comparten el objetivo común de diseñar o usar SRI que le proporcionan información relevante al usuario.
2. Se ha analizado el proceso completo de la recuperación de información para establecer los límites de la recuperación de información en tres ciencias: informática, ciencia de la información y documentación. De esta manera, la informática se ocupa del diseño de SRI, tanto de los componentes de hardware como de software. En la ciencia de la

información se han identificado varias perspectivas y frentes de investigación de los que cabe destacar tres: a) la denominada recuperación de información experimental, que se ocupa del diseño de algoritmos, índices y programas con el objetivo de organizar, almacenar y representar la información de una manera automática y, por consiguiente, constituye la perspectiva informática que se encarga de la entrada en el sistema y del diseño de SRI; b) la recuperación de información en línea, relacionada con el proceso de búsqueda y que constituye el enfoque relacionado con la ciencia de la documentación; y c) la perspectiva cognitiva, centrada en el estudio del comportamiento del usuario y en la comprensión de sus necesidades. Además, tanto en informática como en ciencia de la información se han identificado otras ciencias y disciplinas que contribuyen especialmente al diseño de SRI más eficaces, como la lingüística, las ciencias cognitivas, la lógica y la psicología, entre otras.

3. Se han establecido los límites de la recuperación de información dentro del contexto de la ciencia de la documentación al identificarla como aquella parte del *proceso documental* relacionada con la *búsqueda de información*, desligándola de esta manera de aquellas partes del proceso de recuperación de información relacionadas con la entrada en el sistema, y que corresponden a otras disciplinas de la ciencia de la documentación que se ocupan de la indización, resumen, descripción bibliográfica, representación y organización de la información, tareas y actividades todas ellas realizadas intelectualmente por documentalistas, y que, en modo alguno, coinciden con estas tareas automatizadas más propias, al menos tradicionalmente, de la ciencia de la información debido especialmente a que para ello se requiere de sólidos conocimientos en programación y matemático-estadísticos. En la actualidad los planes de estudio de documentación en las universidades españolas presentan, en general, una formación deficiente en dichas áreas.⁶⁶

A la vista de todas las consideraciones anteriores podemos definir la recuperación de información dentro del contexto de la ciencia de la documentación como *la aplicación del conjunto de técnicas, métodos y actividades para buscar, localizar y recuperar de una manera eficiente en los diversos SRI la información relevante que requiere el usuario, y satisfacer así su necesidad de información*. Definición que es acorde con

66 J. A. Salvador Oliván, "Formación en Recuperación de la Información: Análisis de los cursos y asignaturas en las escuelas de Biblioteconomía y Documentación de Norteamérica y España", en *Documentación de las Ciencias de la Información*, 2003, vol. 25, pp. 189-215.

el concepto de recuperación de información proporcionado por los autores Walker, Harter y Lancaster, cuyos libros de texto se centran de manera especial en la búsqueda de información en bases de datos.

BIBLIOGRAFÍA

- Baeza-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval*, New York: ACM Press, 1999.
- Blair, D. C. *Language and representation in information retrieval*. Amsterdam: Elsevier Science Publishers, 1990.
- Buckland, M. "The landscape of information science: The American Society for Information Science at 62", en *Journal of the American Society for Information Science*, 1999, vol. 50, n° 11, pp. 970-974.
- Bush, V. "As we may think". *Atlantic Monthly*, 1945, 176/1, pp. 101-108. <<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush0.shtml>>
- Caridad Sebastian, M. *Fundamentos Teóricos en Documentación Automatizada: Programa y Bibliografía*. Barcelona: PPU, 1993.
- Denning, P.J. "Computer Science: The discipline", en: A. Ralston y D. Hemmendinger (Eds). *Encyclopedia of Computer Science*, 2000 Edition, London: Nature Publishing Group, 2000. <<http://www.idi.ntnu.no/~ekaterip/dif8916/denning.pdf>>
- Ellis, D. *Progress and problems in Information Retrieval*, London: Library Association Publishing, 1996.
- Harter, S.P. *Online information retrieval: Concepts, principles and techniques*, San Diego: Academic Press, 1986.
- Kochen, M. "Views on the foundations of information science", en Anthony Debons (Ed). *Information Science: Search for identity*. New York: Marcel Dekker, Inc., 1974, pp. 171-177.
- Korfhage R.R. *Information Storage and Retrieval*, New York: John Wiley, 1997.
- Lancaster, F.W. "Sistemas avanzados de recuperación de información", en Wilfrid Lancaster y María Pinto (Coords), *Procesamiento de la información científica*. Madrid: Arco/Libros, 2001.
- _____, Warner, A.J. *Information Retrieval Today*, Arlington, Virginia: Information Resources Press, 1993.
- López Yepes, J. "Introducción", en José López Yepes (Coordinador). *Manual de Información y Documentación*. Madrid: Pirámide, 1996, pp. 21-25.
- _____, *La documentación como disciplina. Teoría e historia. Segunda edición actualizada y ampliada*, Pamplona: EUNSA, 1995.
- Martínez Comeche, J.A. "El proceso informativo-documental", en López Yepes, J. (Coordinador), *Manual de Información y Documentación*. Madrid: Pirámide, 1996, pp. 29-47.
- Meadow, C.T.; Boyce, B.R.; Kraft, D.H. *Text Information Retrieval Systems. Second Edition*, San Diego, California: Academic Press, 1999.

- Mooers, C.N. "The theory of digital handling of non-numerical information and its implications to machine economics", en *Technical Bulletin* No. 48. Cambridge, MA: Zator Co., 1950 (Ponencia presentada en Association for Computing Machinery, Rutgers Univ., New Brunswick, NJ, 1950, March 29).
- Mooers, C.N. "Zatocoding applied to mechanical organization of knowledge", en *American Documentation*, 1951, vol. 2, pp. 20-32.
- Moreiro González, J. A. *Introducción bibliográfica y conceptual al estudio evolutivo de la Documentación*, Barcelona: PPU, 1990.
- _____, *Introducción al estudio de la información y la documentación*, Medellín: Editorial Universidad de Antioquía, 1998.
- Pinto Molina, M. *Análisis documental. Fundamentos y procedimientos* (2ª edición). Madrid: Eudema, 1993.
- Rayward, W.B. "The origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID)", en *Journal of the American Society for Information Science*, 1997, vol. 48, n° 4, pp. 289-300.
- Robins, D. "Interactive information retrieval: Context and basic notions", en *Informing Science*, 2000, vol. 3, n° 2, pp. 57-62.
- Sagredo Fernández, F.; Izquierdo Arroyo, J.Mª. *Concepción lógico-lingüística de la documentación*, Madrid: IBERCOM-Red Comnet de la Unesco, 1983.
- Salton, G.; McGill, M.J. *Introduction to Modern Information Retrieval*, New York: McGraw-Hill, 1983.
- Salvador Olivan, J.A. "Formación en Recuperación de la Información: Análisis de los cursos y asignaturas en las escuelas de Biblioteconomía y Documentación de Norteamérica y España", en *Documentación de las Ciencias de la Información*, 2003, vol. 25, pp. 189-215.
- Saracevic, T. "Information science: origin, evolution and relations", en Perri Vakkari and Blaise Cronin (Eds). *Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives*. London: Taylor Graham, 1992, pp. 5-27.
- Saracevic, T. "Information science", en *Journal of the American Society for Information Science*, 1999, vol. 50, n° 12, pp. 1051-1063.
- _____, "Relevance: A review of and a framework for the thinking on the notion in information science", en Karen Sparck Jones and Peter Willett (Eds). *Readings in Information Retrieval*. San Francisco, California: Morgan Kaufman, 1997, pp. 143-165.
- Schrader, A.M. "In search of a name: information science and its conceptual antecedents", en *Library and Information Science Research*, 1984, vol. 6, n° 3, pp. 227-271.
- Sparck Jones, K.; Willett, P. *Readings in Information Retrieval*, San Francisco: Morgan Kaufman Publishers, Inc., 1997.
- Tague-Sutcliffe, J. "The pragmatics of information retrieval experimentation, revisited", en *Information Processing and Management*, 1992, vol. 28, n° 4, pp. 467-490.

- _____, "The pragmatics of information retrieval experimentation, revisited", en *Information Processing and Management*, 1992, vol. 28, n° 4, pp. 467-490.
- Terrada, M^a L.; Lopez Piñero, J.M^a. "Historia del concepto de Documentación", en *Documentación de las Ciencias de la Información*, 1980, vol. 4, pp. 229-248.
- Vakkari, P. "Library and Information Science: Content and scope", en Johan Olaisen, Erland Munch-Petersen, and Patrick Wilson (Eds), *Information Science. From the development of the discipline to social interaction*, Oslo: Scandinavian University Press, 1996, pp. 169-231.
- Van Rijsbergen, C.J. "A non-classical logic for information retrieval", en Sparck Jones, K.; Willett, P. (Eds.) *Readings in Information Retrieval*. San Francisco: Morgan Kaufman Publishers, Inc., 1997, pp. 268-272.
- _____, Van. *Information retrieval*. 2nd edition. London: Butterworth, 1979.
<<http://www.dcs.gla.ac.uk/Keith/pdf/>
- Walker, G.; Janes, J. *Online Retrieval. A Dialogue of Theory and Practice*. Second Edition, Englewood, Colorado: Libraries Unlimited, Inc., 1999.
- White, H.D.; McCain, K.W. "Visualizing a discipline: An author co-citation analysis of Information Science, 1972-1995", en *Journal of the American Society for Information Science*, 1998, vol. 49, n° 4, pp. 327-355.

