

LA INVESTIGACIÓN SOBRE RECUPERACIÓN DE INFORMACIÓN DESDE LA PERSPECTIVA CENTRADA EN EL USUARIO: MÉTODOS Y VARIABLES

Carmen Caro-Castro*, Lucía Cedeira Serantes*,
Crispulo Travieso Rodríguez*

Resumen: La incorporación del usuario a los estudios sobre recuperación de información ha supuesto el desarrollo de una línea de investigación que centra su atención en cómo se formalizan las necesidades de información y se plantean las búsquedas, cómo interactúa el usuario con el sistema, cómo se valoran los resultados obtenidos y qué características individuales influyen en el proceso. En este estudio exploratorio se analizan 25 trabajos de investigación original que se enmarcan en la perspectiva centrada en el usuario. Se ha establecido una clasificación para las variables cuyas categorías principales son: características de los usuarios, características del sistema, entorno y proceso de búsqueda, y resultados. La coincidencia en la utilización de las técnicas de recogida de datos, métodos de análisis y variables ha servido para comprobar la similitud de las investigaciones y para obtener una representación gráfica de las diferentes orientaciones que se pueden observar entre ellas.

Palabras clave: usuarios, sistemas de recuperación de información, comportamiento de búsqueda, métodos de investigación.

Abstract: User has been included in the research in information retrieval, and this factor has implied a new research approach. This perspective is focused on new issues related to searching process, such as formulation of queries, interaction between user and system, evaluation of the results obtained and influence of some personal characteristics. This exploratory paper examines 25 original research works of the user-centered perspective. A classification of the variables according to the following categories has been established: user characteristics, searching environment and process, and results. The coincidence in data collecting techniques, analysis methods and variables has served for checking the similarity between the research works analysed. Finally, a graphical representation of the different trends observed in these works is presented.

Keywords: users, search behaviour, information retrieval systems, research methods.

1 Introducción

Como señalaban Robertson y Hancock-Beaulieu (1), una aproximación rigurosa a la investigación sobre los sistemas de recuperación de información debería comenzar por la cuestión de dónde situar los límites del sistema. De hecho, las diferencias a la hora de

* Departamento de Biblioteconomía y Documentación. Universidad de Salamanca. Correo-e: ccaro@usal.es.
Recibido: 2-4-02; 2.ª versión: 16-2-03.

emplazarlos han servido para distinguir dos líneas de investigación en esta área: un enfoque basado en el sistema —tradicional o algorítmico— y un enfoque centrado en el usuario (2, 3). Para el enfoque tradicional el «sistema» es el mecanismo que permite almacenar y recuperar la información, la estrategia de búsqueda es el *input* y los documentos recuperados el *output*. Esta delimitación contrasta con la situación que se produce en el mundo real, donde las necesidades de información surgen en contextos concretos y las personas están implicadas en el proceso de búsqueda. La perspectiva centrada en el usuario amplía los límites del «sistema» con la incorporación del componente humano y del entorno en el que se produce la búsqueda. El cambio de enfoque se relaciona con diferentes factores de carácter pragmático y teórico:

- Por una parte, al popularizarse el uso de las redes de información, la recuperación se ha desplazado desde el terreno de los especialistas al de los usuarios finales. Esta nueva población de usuarios, cada vez más numerosa y heterogénea, reclama mecanismos de recuperación que se adapten a sus habilidades y necesidades: herramientas interactivas y amigables que proporcionen acceso a información multimedia y faciliten diferentes tipos de búsqueda (equiparación, exploración, navegación). Para diseñar estos sistemas es necesario saber qué demandan los usuarios y cómo se comportan durante la búsqueda (4).
- Paralelamente, se ha producido un cambio en el concepto de la naturaleza de la recuperación de información. Ya no se considera que el *input* sea una estrategia de búsqueda sino un usuario con una necesidad de información, con estado anómalo de conocimiento (5). En el extremo opuesto, el *output* no es un conjunto de documentos potencialmente relevantes, sino un usuario cuyo estado de conocimiento se ha modificado durante la interacción.

El interés actual por la investigación sobre la interacción usuario/sistema parece favorecer una aproximación de las dos perspectivas. Sin embargo, se trata de un acercamiento complejo, ya que supone la confluencia de enfoques muy diferentes, tanto en sus fundamentos conceptuales como en sus planteamientos metodológicos.

Para la perspectiva tradicional el objetivo primordial de la investigación es mejorar las técnicas de recuperación y los métodos de representación de la información, de manera que se facilite la equiparación entre búsquedas y documentos. Los trabajos de investigación simulan un entorno de búsqueda artificial, controlado, formado por cuatro componentes: un conjunto de documentos, un sistema de almacenamiento y recuperación, un conjunto de temas que se concretan en enunciados de búsqueda y un conjunto de juicios sobre la relevancia de los documentos para esos enunciados.

La eficacia de los sistemas se evalúa en función de las medidas de exhaustividad y precisión basadas en la relevancia temática que se establecieron en Cranfield. Esta evaluación se realiza en un entorno en el que se controlan las variables, para asegurar que los efectos constatados en la investigación se deben a cambios en alguno de los parámetros del sistema. Aunque la metodología que se emplea está bien consolidada, ha sido criticada al menos en dos aspectos: su limitación en tamaño y escala cuando se compara con sistemas comerciales, y la carencia que supone el hecho de no tener en cuenta la contribución humana al proceso de búsqueda.

La crítica sobre tamaño y escala se ha abordado en las Text REtrieval Conferences (TREC) (6). En estos experimentos, coordinados por el National Institute of Standards

and Technology (NIST), diferentes equipos de investigación prueban sus sistemas de recuperación de información con colecciones experimentales que contienen varios gigabytes de datos. Se utilizan temas de búsqueda que intentan reproducir situaciones reales y se establecen juicios sobre la relevancia de los documentos de la colección. Aunque no fue su objetivo original, también han afrontado la segunda crítica, la eliminación del componente humano, introduciendo una línea de trabajo «interactiva» (*TREC Interactive Track*). En esta modalidad, los participantes comprueban la eficacia de diferentes técnicas para mejorar la interacción usuario/sistema siguiendo las directrices generales de las conferencias, de forma que los resultados sean comparables (7).

Del otro lado, la perspectiva centrada en el usuario se inicia con los trabajos de Zweizig y Dervin (8, 9), para quienes la búsqueda y utilización de la información debe analizarse dentro del contexto sociocultural en el que se produce y se resuelve. Básicamente, se trata de conocer las características contextuales e individuales que influyen en la formalización de las necesidades de información, en el desarrollo del proceso de búsqueda y en la valoración de los resultados obtenidos. Bajo este prisma, el rendimiento de los sistemas se evalúa utilizando como criterio la «relevancia basada en la situación», que considera no sólo la coincidencia temática entre búsqueda y documento, sino la utilidad que ésta tiene para el usuario (10).

En realidad, la etiqueta «centrada en el usuario» ha servido para agrupar un conjunto bastante diverso de enfoques que tienen en común la presencia del «factor humano» y la oposición a la perspectiva tradicional (11, 12, 13, 14, 15). Por ejemplo, Sugar (12) señala dos orientaciones: una cognitiva, que estudia cómo se procesa la información, y otra holística, que tiene en cuenta tanto aspectos afectivos y psicomotores como contextuales. Sin embargo, Hert (14) plantea tres enfoques en función del objeto principal de estudio y de la metodología empleada: uno que se centra en la relevancia contextual, otro que analiza la influencia de las características individuales en el comportamiento de búsqueda —que incluye una orientación cognitiva—, y el que examina los cambios que se producen en los objetivos y en las estrategias de los usuarios desde una perspectiva naturalista. Por su parte, Fidel (15) diferencia dos generaciones dentro de los estudios que se reúnen bajo esta denominación: la primera analiza la influencia que tienen las características de los usuarios sobre la interacción; y la segunda trata de identificar patrones de comportamiento en este proceso.

La confusión conceptual y terminológica que ha provocado esta variedad de puntos de vista tiende a simplificarse reconociendo dos corrientes epistemológicas fundamentales: sociológica y cognitiva (16, 17, 18). La primera se centra en el análisis de las influencias «contextuales» —políticas, sociales, institucionales o culturales— sobre el comportamiento de los usuarios de sistemas de recuperación de información. La corriente cognitiva supone el estudio de los aspectos emocionales y cognitivos de esta conducta, que son independientes del contexto en el que se realiza la búsqueda. Para este último enfoque, que surge a partir de los trabajos de De Mey (19), entender la búsqueda de información significa comprender cómo las categorías conceptuales de los usuarios —su modelo del mundo— afectan a la interacción y se modifican durante la misma. Sus planteamientos han tenido tanta importancia en la investigación sobre recuperación de información que autores como Ellis (2), Belkin (20) o Ingwersen (21, 22) consideran que es el «paradigma cognitivo» el que sirve de verdadero contrapunto a la perspectiva tradicional.

En cuanto a los trabajos empíricos, esta corriente también abarca un heterogéneo conjunto de investigaciones cuyo denominador común es el análisis de situaciones de búsqueda-

da en entornos «reales». Para alcanzar este objetivo se ha incrementado el uso de metodologías cualitativas con las que se intenta explicar el comportamiento humano (23). Sin embargo, y a pesar del interés que ha suscitado este tipo de trabajos en la última década, todavía no se ha consolidado un grupo de medidas normalizadas que permitan comparar los resultados de la investigación. La complejidad de la situación se pone de manifiesto en trabajos como los de Fidel y Soergel (24) o Yuan y Meadow (25) que recogen más de un centenar de variables, sistematizándolas en categorías que abarcan: contexto, características del usuario, tutoriales, tema de búsqueda, base de datos, especialista en recuperación de información, sistema de búsqueda, características del proceso y valoración de los resultados.

Siguiendo la línea de estos dos últimos trabajos, hemos realizado un estudio exploratorio que tiene un doble objetivo:

- Recopilar y sistematizar las técnicas de recogida de datos, métodos de análisis y variables utilizados en trabajos de investigación original en los que se ha estudiado el comportamiento de búsqueda de los usuarios de sistemas de recuperación de información.
- Examinar la posibilidad de obtener una representación gráfica de las tendencias de los trabajos analizados, empleando como principio de agrupación la coincidencia en la utilización de los métodos y variables.

2 Metodología

El primer criterio que se utilizó para elegir las publicaciones que se iban a analizar fue la autoría. Como base para realizar la selección se empleó el ranking de los autores más citados en Documentación (1972-1995) elaborado por White y McCain (26). De los 120 que recoge, se seleccionaron aquellos que aparecían en las especialidades de recuperación de información (experimental o en línea), teoría de usuarios y OPAC's. Sumando los «pesos» obtenidos en estos factores, se elaboró una nueva lista que se contrastó y completó con los nombres reseñados por Persson (27) dentro de la línea *soft* en investigación sobre recuperación de información. Finalmente, se incluyó a Spink y Hert, autoras que han colaborado con investigadores que aparecían en el ranking original, además de haber aportado numerosas publicaciones a esta línea en los últimos años.

En cuanto al tipo de trabajo que se ha analizado, se seleccionaron trabajos que presentaban los resultados de investigaciones originales sobre el comportamiento de búsqueda de usuarios finales o de especialistas en recuperación de información. Además, debían incluir una descripción detallada tanto de la metodología como de las variables utilizadas.

Respecto a las fuentes, se eligieron artículos publicados en revistas que se encuentran entre las de mayor factor de impacto en nuestra área, según el ranking *del Journal of Citation Report* (1996-1999): *JASIS*, *Library & Information Science Research*, *Journal of Documentation*, etc. También se ha incluido un trabajo de la sección «interactiva» de las *TREC*, que además es un ejemplo de las investigaciones del Information Interaction Laboratory de la Rutgers School of Communication, Information and Library Studies en el que colabora Belkin (28, 29).

Por último, se aplicó un criterio cronológico: la fecha de publicación debía estar comprendida entre 1980-1999. De tres autores (Borgman, Fidel y Saracevic) se han

seleccionado dos estudios aparecidos con cinco o más años de diferencia, con la intención de comprobar si existían cambios de tendencia a lo largo del periodo analizado.

Los resultados que se presentan son producto del análisis de contenido de 25 artículos (20 trabajos de investigación). Se trata de una muestra que, aunque no sea significativa en términos estadísticos, ha servido para completar una nutrida lista de variables que ha permitido abordar los dos objetivos básicos de este trabajo exploratorio.

Para sistematizar las variables se tomaron como punto de partida las categorizaciones realizadas por Fidel y Soergel (9) y por Yuan y Meadow (10), aunque se fueron modificando a medida que se analizaban los artículos. La heterogeneidad de los estudios empíricos nos ha planteado dificultades, tanto conceptuales como terminológicas, para establecer una clasificación.

En el apartado de la metodología, se optó por diferenciar entre técnicas de recogida de datos y métodos de análisis, división que está relacionada con el desarrollo lógico del proceso de investigación (30). En el grupo de técnicas de recogida de datos, se incluyeron los sistemas utilizados para registrar la información relativa a cualquier persona, objeto o proceso objeto de estudio. En el apartado métodos de análisis se han reseñado los procedimientos empleados para analizar la información original con el objetivo de derivar unos resultados que permitan explicar el fenómeno estudiado.

Respecto al concepto de variable, se ha interpretado en un sentido amplio, considerándose como tal el valor (numérico o no) de cualquier observación o registro realizado por una persona o máquina y el deducido o calculado de alguna manera de las observaciones o registros. Estos valores describen situaciones, personas, sistemas, acciones realizadas y opiniones emitidas (25).

El segundo problema fue elegir la denominación que se iba a emplear para designar una misma metodología, proceso o característica. Entre los términos y expresiones utilizados por los autores, se han elegido los que aparecen en la bibliografía en castellano con mayor frecuencia.

Una vez realizada la clasificación, se elaboró una matriz en la que se indicó la presencia o no de los métodos y variables en cada estudio. A partir de ella se calculó la similitud entre los documentos —utilizando el coeficiente de Dice—, que sirvió de base para obtener la representación gráfica del área de investigación mediante escalamiento multidimensional (MDS). Este método exploratorio permite identificar patrones de asociación en corpus de datos, mediante una representación que facilita la visualización de posibles grupos de objetos. Se ha empleado para establecer relaciones entre fenómenos infoméricos, utilizando diferentes características de la información. Por ejemplo, Kopcsa y Schiebel (31) lo utilizaron para entender la relación entre disciplinas y la organización del espacio intelectual en bases de datos, basándose en la coaparición de palabras o descriptores en los documentos. Ross y Wolfram (32) recurrieron al análisis de cluster y el MDS para derivar relaciones entre los términos de búsqueda de los usuarios de Excite. En el ámbito de la Documentación, White y McCain (26) o Persson (27) lo emplearon para determinar relaciones de especialización a partir de la cocitación de los autores.

3 Resultados

3.1 Técnicas de recogida de datos y métodos de análisis

Ateniéndonos a las limitaciones y definiciones establecidas en el apartado de metodología, el resultado del análisis de los trabajos se sintetiza en la tabla I.

Tabla I
Técnicas de recogida de datos y métodos de análisis

<i>Técnica de recogida de datos</i>	<i>Método de análisis de datos</i>
<ul style="list-style-type: none"> • Cuestionarios • Registro de las transacciones • Grabación de los resultados • Grabación audio de la interacción durante la búsqueda (<i>Protocolos verbales</i>) • Entrevistas • Grupos de discusión • Observación 	<ul style="list-style-type: none"> • Métodos estadísticos <ul style="list-style-type: none"> Métodos descriptivos Métodos inferenciales • Métodos cualitativos <ul style="list-style-type: none"> Análisis de contenido Teoría fundamentada (<i>Grounded theory</i>) Descripción densa

La técnica de recogida de datos más utilizada es el registro de las transacciones usuario-sistema —70%—, seguida de cuestionarios y entrevistas —45% en ambos casos— y de los protocolos verbales —30%—. El porcentaje total supera ampliamente el 100% porque lo habitual es que en una misma investigación se utilicen varias técnicas: únicamente en dos casos se utiliza sólo una y en la mayoría de ellos se combina el registro de las transacciones con los cuestionarios, la entrevista personal o los protocolos verbales.

En cuanto a los métodos de análisis, los estadísticos se emplean en un 85% de los casos: en el 30% sólo se hace una descripción de los resultados y en el 55% se utilizan métodos inferenciales (correlación, chi-cuadrado, regresión, U Man-Whitney y ANOVA). Los métodos cualitativos se aplican en el 55% de las investigaciones, el análisis de contenido de protocolos verbales, entrevistas o transacciones en un 45%, la teoría fundamentada en el 20% y la descripción densa en el 15%.

Como se deduce de estos datos, es frecuente que en una misma investigación se utilicen varios métodos de análisis de los asignados a una misma clase. Además, en el 35% de los casos se utilizan simultáneamente métodos cualitativos y cuantitativos lo que demuestra la tendencia a la integración de las dos orientaciones metodológicas.

3.2 Variables

El conjunto de las variables utilizadas se ha dividido en cinco categorías: 1) usuarios, 2) sistema, 3) entorno de la búsqueda, 4) proceso de búsqueda y 5) resultados.

Las variables relacionadas con los *usuarios* representan un conjunto de características que se utilizan para comprobar su influencia en el comportamiento de búsqueda y en la valoración que se hace de los resultados. El objetivo es crear modelos de usuarios basados en las diferencias individuales para poder diseñar sistemas que se adapten a la conducta de grupos con atributos diferentes.

La variable que se ha utilizado en más investigaciones es la experiencia —40%— seguida a distancia de las características demográficas —20%—. Por debajo —15%— se encuentran las características personales (medidas por medio de algún test normalizado), el nivel profesional y el conocimiento del área temática. Con menor frecuencia —10%— se ha tenido en cuenta la formación, los estados emocionales o afectivos y los modelos mentales, entendidos como la idea que el usuario tiene sobre las características, estructura, contenido y funcionamiento del sistema.

Tabla II
Características de los usuarios y del sistema

<i>Características de los usuarios</i>	
<ul style="list-style-type: none">• Características demográficas<ul style="list-style-type: none">• Género• Edad• Formación<ul style="list-style-type: none">• Titulación o formación principal• Nivel• Nivel profesional• Experiencia<ul style="list-style-type: none">• Con el SRI• Con la base de datos• Con los ordenadores en general• Con la búsqueda de información	<ul style="list-style-type: none">• Conocimiento de la materia y del área temática• Modelos mentales• Características personales<ul style="list-style-type: none">• Inteligencia• Estilo de aprendizaje• Habilidad para el razonamiento simbólico• Habilidad lingüística• Creatividad• Estados emocionales o afectivos
<i>Características del sistema</i>	
<ul style="list-style-type: none">• Base de datos<ul style="list-style-type: none">• Campo temático• Número de registros• Técnica de recuperación de información	<ul style="list-style-type: none">• Interfaz• Organización de la información• Tipos de índices• Tutorial

Las variables relacionadas con el *sistema* se utilizan en aquellos estudios que comparan diferentes sistemas de recuperación, el mismo con diferentes bases de datos o la eficacia de distintas interfaces. Habitualmente, se consideran variables independientes y se relacionan con el rendimiento que obtienen los usuarios y con las características del proceso de búsqueda. Las que más interesan son las características de la base de datos y las técnicas de recuperación de información —30% en ambos casos—. La forma de organizar la información se examina en el 15% de los estudios, mientras que la interfaz, los tipos de índices y las características del tutorial en línea sólo se estudian en el 10%.

Las vinculadas al *entorno de búsqueda* se han subdividido en dos grupos: contexto de búsqueda y definición del problema.

Las primeras sirven para estimar la influencia del entorno en el que se enmarca la búsqueda desde un punto de vista pragmático: para qué se busca la información —25%—, en qué fase se encuentra el proyecto para el que se va a utilizar —15%—, a qué campo temático pertenece —10%—, si se trata de un entorno teórico, práctico o general —10%— y cuánta información se presupone que existe sobre el tema —10%.

En las segundas, se aborda el proceso de transformación de la necesidad de información en un enunciado que la sintetice: nivel de definición —25%—, expresión —15%— y número de búsquedas —25%. Este grupo se utiliza para comprobar cómo influyen estos factores en el proceso de búsqueda y en la valoración de los resultados y cómo dependen de las características de los usuarios.

Tabla III
Entorno y proceso de búsqueda

<i>Entorno de búsqueda</i>	
<i>Contexto de la búsqueda de información</i>	<i>Definición del problema</i>
<ul style="list-style-type: none"> • Entorno de trabajo en el que se inserta la búsqueda (área temática, teórico, práctico, etc.) • Finalidad de la búsqueda • Fase de desarrollo del proyecto para el que se busca información • Tiempo que se ha previsto invertir en la búsqueda • Presuposiciones (cantidad de información que se cree que existe sobre el tema de búsqueda) 	<ul style="list-style-type: none"> • Nivel de definición de la necesidad de información • Expresión de la necesidad de información <ul style="list-style-type: none"> Claridad Especificidad Complejidad • Dificultad • Número de búsquedas por necesidad de información expresada
<i>Proceso de búsqueda</i>	
<ul style="list-style-type: none"> • Puntos de acceso utilizados • Tipo de punto de acceso • Descriptores utilizados <ul style="list-style-type: none"> Número total Tipo de descriptores • Fuente de los descriptores <ul style="list-style-type: none"> Estructuras de conocimiento del sistema Otros documentos Retroalimentación automática • Combinaciones de términos de búsqueda <ul style="list-style-type: none"> Número de combinaciones Tipo de operador utilizado • Órdenes y opciones del sistema utilizadas • Movimientos <ul style="list-style-type: none"> Número de movimientos Tipo de movimiento 	<ul style="list-style-type: none"> • Ciclos <ul style="list-style-type: none"> Número de ciclos Tipo de ciclo • Visualización de registros o documentos <ul style="list-style-type: none"> Número de registros visualizados en cada ciclo de búsqueda Número total de registros visualizados • Errores <ul style="list-style-type: none"> Número de errores Tipo de error Circunstancias del tipo de error • Tiempo empleado en la sesión <ul style="list-style-type: none"> Tiempo empleado en la resolución de problemas Tiempo empleado en ejecutar la búsqueda • Reacciones y emociones del usuario durante la búsqueda • Modificación del objetivo inicial de la búsqueda

Las agrupadas bajo el epígrafe *proceso de búsqueda* son variables relacionadas con la interacción directa del usuario con el sistema de recuperación y tienen como objetivo reflejar las acciones realizadas durante esta fase y entender qué razones las justifican, con la finalidad de obtener patrones del proceso. Se utilizan como variables dependientes de las características del usuario o del sistema, e independientes cuando se analiza cómo varía el rendimiento en función de la estrategia de búsqueda.

El número de puntos de acceso —10%— o de descriptores empleados —50%— se considera un índice del grado de interacción de la persona que busca información y defi-

ne la complejidad de la búsqueda. La combinación de términos —15%— y la utilización de las órdenes y opciones del sistema —40%— origina diferentes tipos de movimientos —30%— y ciclos de búsqueda —45%—. La clasificación de los movimientos —20%— y de los ciclos —15%— se hace tanto en función de las acciones realizadas (selección y combinación de términos, visualización de registros o índices, etc.) como de lo que se pretende obtener con ellas (limitar la búsqueda, ampliarla, entender el funcionamiento del sistema, etc.).

Otros factores considerados para caracterizar el proceso de búsqueda son la visualización de registros o documentos —25%—, los errores cometidos —20%— y el tiempo empleado —20%—. En los trabajos de corte más cualitativo se valoran también aspectos como los estados emocionales y las reacciones del usuario —10%— o la modificación del objetivo de la búsqueda en función de los registros visualizados o de los primeros resultados obtenidos —15%—.

Finalmente, queda el grupo de variables relacionadas con los *resultados*. Las medidas de rendimiento que se han utilizado combinan los dos puntos de vista en la evaluación de sistemas de recuperación de información. Por una parte se mantienen las medidas tradicionales, exhaustividad y precisión, basadas en la relevancia temática —30%—. Desde otra perspectiva, la idea de «búsqueda en situación» introduce un principio de incertidumbre en la valoración de la relevancia porque pasa a considerarse una medida dependiente del contexto. En lugar de relevancia, algunos autores emplean el término «utilidad» —25%— para denominar el criterio que permite valorar la integridad y la exactitud de los resultados. Las medidas de eficiencia tratan de evaluar, con carácter más o menos objetivo, la relación entre el coste de la búsqueda (económico 10% o tiempo 10%) y el resultado obtenido.

Tabla IV
Rendimiento

<i>Rendimiento</i>	
<ul style="list-style-type: none">• Número de registros recuperados <i>Eficacia</i> <ul style="list-style-type: none">• Juicios de relevancia por registro<ul style="list-style-type: none">Relevancia temáticaRelevancia contextual (<i>Utilidad</i>)• Juicios de relevancia por búsqueda<ul style="list-style-type: none">ExhaustividadIntegridad de los resultadosPrecisiónExactitud de los resultados	<i>Eficiencia</i> <ul style="list-style-type: none">• Coste de la búsqueda• Valoración subjetiva del coste de la búsqueda en función de los resultados• Valor de los resultados de búsqueda en función del tiempo empleado• Valor de la búsqueda en su conjunto

3.2 Representación gráfica

Para mostrar gráficamente las relaciones entre los trabajos analizados se eligió la representación tridimensional obtenida mediante MDS (stress de Kruskal .14) porque se ajustaba mejor a los datos analizados que la bidimensional (stress .23) y resultaba más fácil de visualizar que la de cuatro dimensiones, sin que ésta última añadiera ninguna información sustancial. En cualquier caso, para facilitar la visualización, se presenta la misma información en dos gráficos de dos dimensiones.

Figura 1
Mapa de relaciones en función de técnicas, métodos y variables

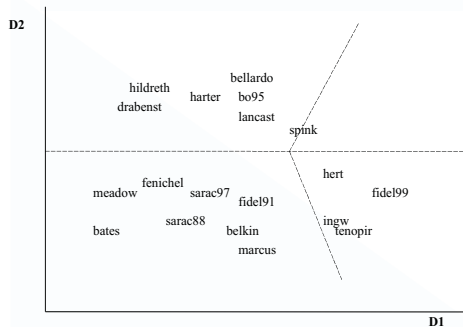
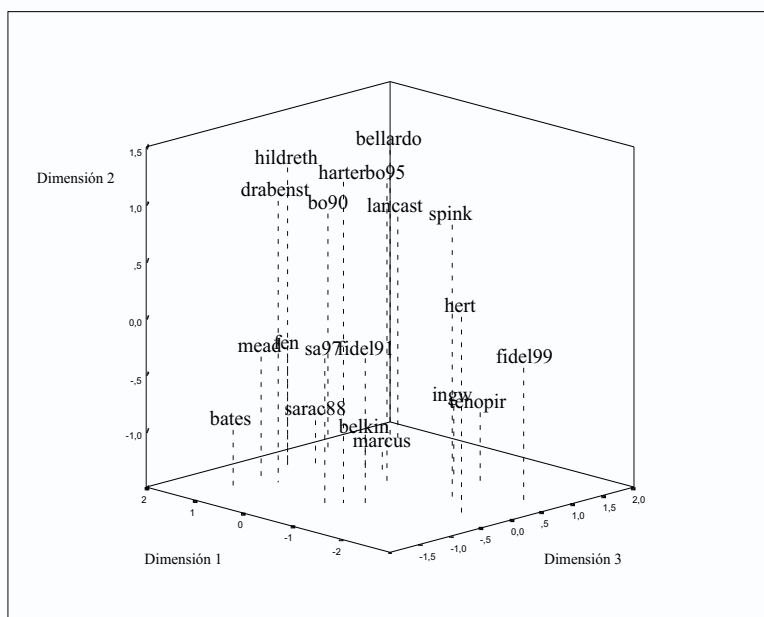


Fig. 1.1

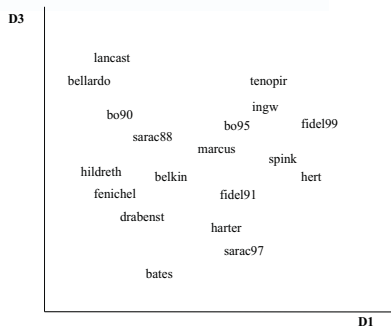


Fig. 1.2

Como puede apreciarse, no se obtienen agrupaciones muy contrastadas y las concentraciones más claras se producen en función del eje vertical (dimensión 2). En la parte superior se reúnen los trabajos cuyo interés de investigación fundamental es la influencia de las características de los usuarios o de diferentes sistemas de recuperación de información en el desarrollo del proceso de búsqueda y en el rendimiento final [3, 4, 5, 6, 10, 12, 14, 19] (en el anexo I se indica entre corchetes la etiqueta que identifica los trabajos en el gráfico). Como rasgos del proceso se consideran los descriptores seleccionados, el tiempo invertido, los errores cometidos y las órdenes o movimientos realizados. Además, en dos trabajos se evalúa la influencia de la formación proporcionada

por diferentes tipos de tutorial, pero no se utilizan variables que permitan estimar la influencia del contexto de búsqueda, y sólo Spink [19] valora la definición del problema.

La información se recoge especialmente mediante cuestionarios aunque también se emplean la entrevista y la grabación de las transacciones o los protocolos verbales y grupos de discusión. Todos utilizan métodos de análisis estadísticos que en cinco ocasiones es inferencial [3, 4, 6, 10, 12] y en otras cinco se combina con algún tipo de análisis cualitativo [5, 6, 10, 14, 19].

Los autores que se concentran en la parte inferior se interesan por la caracterización del proceso de búsqueda, empleando todas las variables que permiten representarlo [1, 2, 7, 8, 9, 11, 13, 15, 16, 17, 18, 20]. Se considera que el proceso puede variar por factores relacionados con el sistema (interfaz, base de datos, tipo de índices), con alguna característica de los usuarios (fundamentalmente la experiencia) o con el entorno de la búsqueda (contexto y definición del problema) y también se examina cómo pueden influir las características del proceso en los resultados.

Las técnicas fundamentales de recogida de datos son las transacciones seguidas de las entrevistas, protocolos verbales, cuestionarios, grupos de discusión y observación. En siete investigaciones se utilizan métodos cualitativos [2, 8, 9, 11, 13, 18, 20] mientras que todos emplean métodos estadísticos —en seis casos inferenciales [2, 8, 13, 16, 17, 18]— excepto Tenopir [20], Hert [11] y Fidel [9].

Estos tres últimos trabajos junto al de Ingwersen [13] forman el subgrupo que aparece a la derecha de este conjunto de autores que se interesa básicamente por el proceso de búsqueda. En ellos se intenta comprenderlo desde una perspectiva holística, teniendo en cuenta sobre todo la influencia del entorno y la definición del problema, los cambios que se producen en el objetivo de la búsqueda a lo largo del proceso y las reacciones emocionales del usuario durante la búsqueda. Esta subdivisión se explica fundamentalmente por motivos metodológicos (dimensión 1): se utilizan pocas variables y métodos cualitativos (teoría fundamentada, descripción etnográfica) para analizar la información proporcionada por transacciones, protocolos verbales, entrevistas y observación. Estas mismas razones justifican la aproximación al grupo, en la parte superior, del trabajo de Spink [19].

En cuanto a la distribución en profundidad (dimensión 3) se explica por la mayor o menor presencia de variables relacionadas con la valoración de los resultados y con los usuarios. Sin embargo, no genera grupos diferenciables, aunque mejora el ajuste de la representación.

Por último, los desplazamientos producidos entre los trabajos de un mismo autor, son poco significativos ya que en los tres casos se mantienen básicamente dentro del mismo grupo. El movimiento más amplio quizá sea del de Fidel y se justifica esencialmente por un cambio en las técnicas de recogida de datos y métodos de análisis.

4 Conclusiones

Los trabajos analizados confirman la heterogeneidad de las metodologías y variables utilizadas en el marco de la línea de investigación centrada en el usuario. La falta de consistencia se produce tanto por la variedad de aspectos del proceso de búsqueda que se estudian como por la falta de acuerdo a la hora de definirlos y denominarlos. En cual-

quier caso, esta situación hace que la sistematización y la comparación de los trabajos sea muy complicada.

En el terreno de las metodologías resulta evidente la combinación de técnicas y métodos que tiene como objetivo alcanzar una imagen más completa y una comprensión más profunda del fenómeno que se estudia. Esta fusión, que demuestra la continuidad entre lo cuantitativo y cualitativo, puede realizarse utilizando estrategias de complementación, combinación o triangulación (33).

Aunque no se puede vincular con claridad esta línea de investigación con una metodología concreta, la coincidencia temática provoca cierta homogeneidad a la hora de utilizar las técnicas de recogida de datos y los métodos de análisis. El interés por obtener información sobre la interacción usuario-sistema de la forma menos obtrusiva posible justifica el predominio del análisis transaccional que aporta información susceptible de ser analizada cuantitativa o cuantitativamente (34). La utilización de técnicas y métodos cualitativos es imprescindible cuando el objetivo no es sólo describir el proceso sino entender las causas y las consecuencias de las acciones realizadas durante el mismo (23).

En relación con las variables, no sólo llama la atención su gran variedad, sino la falta de coincidencia tanto en su definición como en su utilización. La menos uniforme de las categorías establecidas es la del proceso de búsqueda y la más difícil de caracterizar la del entorno. En la primera, los aspectos estudiados responden a criterios de análisis tan distintos, y tratan de abarcar opciones de búsqueda de sistemas tan diferentes, que son difíciles de etiquetar bajo un mismo nombre de variable: ¿se incluyen en la misma categoría los ciclos o movimientos que se establecen en función de los objetivos y los que se justifican por la metodología?, ¿sirven las mismas variables para representar las búsquedas por equiparación exacta y las exploratorias? Respecto al entorno, la dificultad para caracterizarlo se relaciona con la amplitud de situaciones que se intenta abarcar, con su carácter subjetivo y con que es la categoría en la que más han cambiado los criterios de la investigación.

La falta de un conjunto estable de variables ha servido como argumento para cuestionar la solidez de esta línea de investigación (34). Por eso, quizá la conclusión más importante de esta revisión sea la conveniencia de homogeneizarlas y reducir su número para facilitar tanto el contraste de los resultados como la creación de un marco conceptual consistente.

A pesar de las limitaciones que supone la diversidad que acabamos de indicar, el resultado gráfico obtenido concuerda con algunas de las líneas de trabajo señaladas dentro de la perspectiva centrada en el usuario. Una valora fundamentalmente la influencia de las características de los usuarios y de la interacción con diferentes sistemas en los resultados obtenidos. Otra se preocupa por el análisis y caracterización del proceso de búsqueda y cómo le afectan los atributos del sistema, de la búsqueda o de los usuarios. Dentro de esta última se puede considerar una subdivisión de carácter más cualitativo, holístico-naturalista y centrada en aspectos relacionados con el entorno de búsqueda y la representación de la necesidad de información. Estas dos últimas ramificaciones concuerdan con las que plantean Hert (14) y Sugar (12).

Es cierto que esta heterogeneidad invita a preguntarse cómo cambiaría la representación gráfica si se analizaran más trabajos. Sin embargo, pensamos que el número de variables que se ha tenido en cuenta ha servido para crear un entramado que permite perfilar la tendencia de los trabajos de investigación producidos en esta área, aunque para confirmar su validez sería necesario ampliar el ámbito de este estudio exploratorio

Bibliografía

1. ROBERTSON, S. E.; HANCOCK-BEAULIEU M. M.. On the evaluation of IR systems. *Information Processing & Management*, 1992, 28(4), p. 457-466.
2. ELLIS, D. The physical and cognitive paradigms in information retrieval research. *Journal of Documentation*, 1992, 48(1), p. 45-64.
3. HARTER, S. P.; HERT, C. A. Evaluation of information retrieval systems: approaches, issues, and methods. *Annual Review of Information Science and Technology*, 1997, 32, p. 3-93.
4. MIRA: evaluation frameworks for interactive multimedia information retrieval applications. [en línea]. Disponible en: <http://www.dcs.gla.ac.uk/mira> (Consulta: 12 febrero 2003).
5. BELKIN, N. J.; ODDY R. N.; BROOKS, H. M. ASK for information retrieval. Part I: Background and theory. *Journal of Documentation*, 1982, 38(2), p. 61-71.
— ASK for information retrieval. Part II: Results of a desing study. *Journal of Documentation*, 1982, 38(3), p. 145-164.
6. *Text REtrieval Conference* [en línea]. Disponible en: <http://trec.nist.gov/> (Consulta: 12 febrero 2003)
7. BEAULIEU, M.; ROBERTSON, S.E.; RASMUSSEN, E.M. Evaluating interactive systems in TREC. *Journal of the American Society for Information Science*, 1996, 47, p. 85-94.
8. ZWEIZIG, D.L. With our eye on the user: need research for information and referral in the public library. *Drexel Library Quarterly*, 1977, 12, p. 48-58.
9. ZWEIZIG, D.L.; DERVIN, B. Public library use, users, uses: advances in knowledge of the characteristics and needs of the adult clientele of american public libraries. En: Voigt, M.J.; Harris, M.K. (eds.). *Advances in Librarianship*. New York: Academic Press, 1977, p. 231-255.
10. NAHL, D. The user-centered revolution. En: Kent, A.; Hall, C.M. (eds.) *Encyclopedia of Library and Information Science*. A. New York: Marcel Dekker, 1998. v. 62 (supl. 25), p. 313-371.
11. DALRYMPLE, P.W. A quarter century of user-centered study: the impact of Zweizig and Dervin on LIS research. *Library & Information Science Research*, 2001, 23(2), p. 155-165.
12. SUGAR, W. User-centered perspective of information retrieval research and analysis methods. En: Williams, M.E. (ed.) *Annual Review of Information Science and Technology*. Medford: Learned Information, 1995, v. 30, p. 77-109.
13. ALLEN, B. L. *Information task: toward a user-centered approach to information systems*. San Diego: Academic Press, 1996.
14. HERT, C. A. *Understanding information retrieval interactions: theoretical and practical implications*. Greenwich, Ablex, 1997.
15. FIDEL, R. The user centered approach: how we got here. En: Wheeler, W.J (Ed.) *Saving the time of the library user through subject access innovation: papers in honor of Pauline Atherton Cochrane*. W. J. Champain, Il: Graduate School of Library and Information Science, 2000.
16. SONNENWALD, D. H.; IIVONEN, M. An integrated human information behavior research framework for information studies. *Library & Information Science Research*, 1999, 21(4), p. 429-457.
17. PETTIGREW, K. E.; FIDEL, R.; BRUCE, H. Conceptual frameworks in information behavior. En: Williams, M.E. (ed.) *Annual Review of Information Science and Technology*. Medford: Learned Information, 2001, v. 35, p. 43-77.
18. FERNÁNDEZ MOLINA, J. C. ; MOYA ANEGÓN, F. Perspectivas epistemológicas «humanas» en la documentación. *Revista Española de Documentación Científica*, 2002, 25(3), p. 241-253.
19. DE MEY, M (ed.). *International Workshop on the Cognitive Viewpoint, 1977 March 24-26, Ghent, Belgium*. Ghent, Belgium: University of Ghent, 1977.
20. BELKIN, N. The cognitive viewpoint in Information Science. *Journal of Information Science*, 1990, 16(1), p. 11-16.

21. INGWERSEN, P. *Information retrieval interaction*. London: Taylor Graham, 1992.
22. INGWERSEN, P. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 1996, 52(1), p. 3-50.
23. FIDEL, R. Qualitative methods in information retrieval research. *Library & Information Science Research*, 1993, 15(3), p. 219-247.
24. FIDEL, R.; SOERGEL, D. Factors affecting online bibliographic retrieval: a conceptual framework for research. *Journal of the American Society for Information Science*, 1983, 34, p. 163-180.
25. YUAN, W.; MEADOW, C. T. A study of the use of variables in information retrieval user studies. *Journal of the American Society for Information Science*, 1999, 50(2), p. 140-150.
26. WHITE, H. D.; MCCAIN, K. W. Visualizing a discipline: an author co-citation analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science*, 1998, 49(4), p. 327-355.
27. PERSSON, O. The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science*, 1994, 45(1), p. 31-38.
28. OVER, P. The TREC interactive track: an annotated bibliography. *Information Processing & Management*, 2001, 37(3), p. 369-381.
29. BELKIN, N. J. An overview of results from Rutgers' investigation of interactive information retrieval. En: *Visualizing subject access for Twenty first Century information resources: proceedings of the 1997 Clinic on Library Applications of Data Processing, 2-4 March 1997*. P. A. Cochrane & E. H. Johnson (Eds.) Urbana-Champaign, Illinois: Illinois University at Urbana-Champaign, Graduate School of Library and Information Science, 1998, p. 45-62.
30. McCLURE, C. User-based data collection techniques and strategies for evaluating networked information services. *Library Trends*, 1994, 42(4), p. 591-607.
31. KOPCSA, A.; SCHIEBEL, E. Science and technology mapping: a new interaction model for representing multidimensional relationships. *Journal of the American Society for Information Science*, 1998, 49(1), p. 7-17.
32. ROSS, N.; WOLFRAM, D. End user searching on the Internet: an analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science*, 2000, 51(10), p. 949-958.
33. BORREGO HUERTA, A. *Metodología cualitativa de investigación en Biblioteconomía y Documentación: 1993-1999*. Universidad de Salamanca, 2001. Tesis doctoral inédita.
34. BORGMAN, C. L.; HIRSH, S. G.; HILLER, J. Rethinking online monitoring methods for information retrieval systems: from search product to search process. *Journal of the American Society for Information Science*, 1996, 47(7), p. 568-583.
35. FIDEL, R. What is missing in research about online searching behavior. *The Canadian Journal of Information Science*, 1987, 12(3/4), p. 54-61.

Anexo I. Artículos analizados

- [1] [BATES] BATES, M. J.; WILDE, D. N.; SIEGFRIED, S. An analysis of search terminology used by humanities scholars: the Getty Online Searching Project report number 1. *Library Quarterly*, 1993, 63, p. 1-39.
- [BATES] BATES, M. J.; WILDE, D. N.; SIEGFRIED, S. A profile of end-user searching behavior by humanities scholars: the Getty online searching project report nº 2. *Journal of the American Society for Information Science*, 1993, 44(5), p. 273-277.
- [2] [BELKIN] KOENEMANN, J.; QUATRAIN, R.; BELKIN, N. New tools and old habits: the interactive searching behavior of expert online searches using INQUERY. En: *TREC-3: Proceedings of the Third Text Retrieval Conference*. D. Harman. Whashington: GPO, 1995, p. 145-177.

- [3] [BELLARDO] BELLARDO, T. An investigation of online searcher traits and their relationship to search outcome. *Journal of the American Society for Information Science*, 1985, 36(4), p. 241-250.
- [4] [BO90] SULLIVAN, M. V.; BORGMAN, C. L.; WIPPEM, D. End-users, mediated searches, and front-end assistance programs on DIALOG: a comparison of learning, performance, and satisfaction. *Journal of the American Society for Information Science*, 1990, 41(1), p. 27-42.
- [5] [BO95] BORGMAN, C. L.; HIRSH, S. G.; WALTER, V. A. Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog Project. *Journal of the American Society for Information Science*, 1995, 46(9), p. 663-684.
- [6] [DRABENST] DRABENSTOTT, K. M.; WELLER, M. S. Failure analysis of subject searches in a test of a new design for subject access to online catalogs. *Journal of the American Society for Information Science*, 1996, 47(7), p. 519-537.
- [7] [FEN] FENICHEL, C. Online searching: measures that discriminate among users with different types of experience. *Journal of the American Society for Information Science*, 1981, 23, p. 23-32.
- [8] [FIDEL91] FIDEL, R. Searches selection of search keys. I The selection routine. *Journal of the American Society for Information Science*, 1991, 42(7), p. 490-500.
[FIDEL91] FIDEL, R. Searches selection of search keys. II Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science*, 1991, 42(7), p. 501-514.
[FIDEL91] FIDEL, R. Searches selection of search keys. III Searching styles. *Journal of the American Society for Information Science*, 1991, 42(7), p. 515-527.
- [9] [FIDEL99] FIDEL, R.; et al. A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 1999, 50(1), p. 24-37.
- [10] [HARTER] HARTER, S. P.; CHENG, Y. R. Colinked descriptors: improving vocabulary selection for end-user searching. *Journal of the American Society for Information Science*, 1996, 47(4), p. 311-325.
- [11] [HERT] HERT, C. A. User goals on an online public access catalog. *Journal of the American Society for Information Science*, 1996, 47(7), p. 504-518.
- [12] [HILDRETH] HILDRETH, C. R. The use and understanding of keyword searching in a university online catalog. *Information Technology and Libraries*, 1997, 16(2), p. 52-62.
- [13] [INGW] BORLUND, P.; INGWERSEN, P. The development of a method for the evaluation of interactive Information Retrieval Systems. *Journal of Documentation*, 1997, 53(3), p. 225-250.
- [14] [LANCAST] LANCASTER, F. W.; et. al. Searching databases on CD-ROM: comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries. *RQ*, 1994, 33(3), p. 370-386.
- [15] [MARCUS] MARCUS, R. S. An experimental comparison of the effectiveness of computers and humans as search intermediaries. *Journal of the American Society for Information Science*, 1983, 34(6), p. 381-404.
- [16] [MEAD] MEADOW, C. T.; WANG, J.; YUAN, W. A study of user performance and attitudes with information retrieval interfaces. *Journal of the American Society for Information Science*, 1995, 46(7), p. 490-505.
- [17] [SARAC88] SARACEVIC, T.; KANTOR, P.; CHAMIS, A. Y.; TRIVINSON, D. A study of information seeking and retrieving. I: Background and methodology. *Journal of the American Society for Information Science*, 1988, 39, p. 161-176.
[SARAC88] SARACEVIC, T.; KANTOR, P. A study of information seeking and retrieving. II: Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 1988, 39, p. 177-196.

- [SARAC88] SARACEVIC, T.; KANTOR, P. A study of information seeking and retrieving. III: Searches, searches and overlap. *Journal of the American Society for Information Science*, 1988, 39, p. 197-216.
- [18] [SA97] SPINK, A.; SARACEVIC, T. Interaction information retrieval: selection and effectiveness of search terms. *Journal of the American Society for Information Science*, 1997, 48(8), p. 741-761.
- [19] [SPINK] SPINK, A. Multiple search sessions model of end-user behavior: an exploratory study. *Journal of the American Society for Information Science*, 1996, 47(8), p. 603-609.
- [20] [TENOPIR] TENOPIR, C.; DAHL-JAKOBOVITS, D.; HOWARD, D. L. Strategies and assessments online: novices' experience. *Library and Information Science Research*, 1991, 13(3), p. 237-266.