

ARTÍCULOS

Documentación

BOLETÍN DE LA ANABAD XLVIII (1998), NÚM. 1, ENERO-MARZO. MADRID ISSN 0210-4164

El control de la calidad en bases de datos de Ciencias Sociales

Ana EXTREMEÑO
Purificación MOSCOSO*

1. INTRODUCCIÓN

Desde el nacimiento de la industria de la información, el control de la calidad de las bases de datos documentales ha sido motivo de interés y preocupación. Expertos y teóricos en esta materia han abordado la evaluación de estas fuentes electrónicas desde distintas perspectivas.

En un principio solía tratarse de un análisis básicamente cuantitativo, que se limitaba al mero recuento de datos de diversa índole. Así, el núme-

* Facultad de Documentación. Universidad de Alcalá.

ro de registros, o el de revistas o autores recogidos, por ejemplo, eran las constantes que servían para analizar la calidad de una base de datos. El progresivo aumento, tanto de los servicios de información como de las propias bases de datos, hizo necesario no limitar este proceso de evaluación a una simple cuantificación. De esta forma, comenzaron a llevarse a cabo estudios cualitativos con la finalidad de analizar si la información recogida se adecuaba a las necesidades y demandas de los diferentes tipos de usuarios. La evaluación de cuestiones de carácter técnico y la de los propios servicios comerciales empezaron, igualmente, a incluirse en los estudios sobre control de calidad.

Desde principios de la década pasada, la preocupación por el control de la calidad de estas fuentes electrónicas resurge de forma significativa, hecho que se pone de manifiesto a través de las publicaciones dedicadas a este tema. No hay que olvidar, además, que Internet ha desempeñado un papel fundamental, ya que ha favorecido la proliferación de estos recursos así como el acceso a la información que proporcionan.

Una gran parte de los estudios realizados hasta la fecha se han centrado en la evaluación de productos concretos. Otra parte, menor pero significativa, se ha ocupado del análisis de los aspectos que deben tenerse en cuenta a la hora de evaluar bases de datos documentales, así como de establecer metodologías adecuadas a este proceso mediante el uso de distintos indicadores.

Nuestro trabajo se centra en el control de calidad de las bases de datos en Ciencias Sociales, desde un punto de vista global, si bien hemos dado prioridad a los aspectos relacionados con su contenido. No hemos hecho distinción entre las bases de datos accesibles en línea y las almacenadas en CD-ROM, por lo que el análisis llevado a cabo es extrapolable a ambos tipos de soporte.

2. LA CALIDAD COMO CONCEPTO Y OBJETO DE ANÁLISIS

En las distintas definiciones donde se aborda el concepto de calidad la satisfacción del usuario o cliente es una constante. Resolver y satisfacer las expectativas de los usuarios es la razón de ser fundamental en el control de calidad de productos o servicios. Así, la norma ISO 8402 la define como «la totalidad de prestaciones y características de un producto o servicio cuya principal misión es satisfacer necesidades reales o implícitas».

Todas las series 9000 de ISO se refieren a los sistemas de gestión de la calidad, que recogen las instrucciones, pautas y técnicas para la organización y estructuración del proceso, así como el grado de implicación y responsabilidad de los distintos participantes.

Llegar a una definición universal del concepto de calidad, o determinar medidas absolutas para su gestión no es una tarea fácil, ya que el contexto de aplicación influye, decisivamente, en las distintas variables que entran en juego. Asimismo, no podemos obviar que en este proceso influye la perspectiva desde la cual se evalúa, así como el punto de vista de quienes hacen uso del servicio o del producto, lo que da lugar a un cierto grado de subjetividad del que es difícil desvincularse. Existen metodologías que tratan de eludir cualquier resquicio posible de subjetividad, mediante el diseño de tests apriorísticos que permitan medir, de una forma objetiva, la idoneidad de productos y servicios. Sin embargo, como constata Pacios refiriéndose a los sistemas de gestión de calidad total, «para que esa adecuación se dé es necesario contar con las personas que reciben el producto o servicio, cuyo criterio estará en función del uso, valor y utilidad que conceda el mismo»¹.

El concepto de calidad suele asociarse con Japón, aunque actualmente sus máximos entusiastas se encuentran también en Estados Unidos y, recientemente, en Europa. La «calidad», como sistema de gestión susceptible de aplicarse en diversos contextos, surge en Estados Unidos, en el sector industrial, en la década de los años cuarenta. Sin embargo, es Japón el país que más contribuyó a su desarrollo al finalizar la II Guerra Mundial. Desde entonces, hasta la actualidad, el concepto de calidad ha pasado por una serie de fases que han determinado distintas formas de entenderla y aplicarla.

Durante los años cincuenta el control de calidad se asociaba a la mera inspección de los productos. En los sesenta comenzaron a controlarse los procesos de puesta en funcionamiento, y en los setenta este control se refiere, sobre todo, a los sistemas y al procedimiento. En la década de los ochenta el tema del control de calidad se convierte en todo un modelo de gestión, que se aplica a cada una de las etapas y fases del proceso de generación de un producto. Este enfoque da lugar, en los años noventa, a la llamada gestión de calidad total, o, en su acrónimo anglosajón TQM (Total Quality Management).

De esta forma, podríamos resumir que durante los años cincuenta la calidad se considera un lujo, en los sesenta un coste adicional, en los setenta un instrumento de venta, y en los ochenta un elemento diferenciador. En los noventa, el control de calidad es necesario para la supervivencia del producto, debido, entre otras razones, a la competitividad del mercado.

¹ PACIOS LOZANO, A. R. Sistemas de calidad total en bibliotecas y unidades de información. *Boletín ANABAD*, 1995, n 2, p. 107-116, p. 109.

3. LAS BASES DE DATOS DOCUMENTALES COMO PRODUCTOS Y SERVICIOS SUSCEPTIBLES DE EVALUARSE

El control de calidad, aplicado a la gestión de las bases de datos documentales, no cobra protagonismo hasta la década de los ochenta. Sin embargo, en un periodo relativamente corto, se ha convertido en una cuestión prioritaria en todos los sectores implicados en la gestión y producción de información. Dos hechos, ampliamente constatados, dan prueba de ello: el aumento de la bibliografía al respecto², y el nacimiento y proliferación de los llamados «grupos de calidad»³.

Las bases de datos documentales son el producto resultante y tangible de los servicios que ofertan sus distribuidores o proveedores. La calidad de los servicios añadidos al producto va a influir, necesariamente, en la calidad final del producto.

La filosofía que subyace al control de su calidad se engloba dentro de la gestión de la calidad de los servicios de información, por lo que implica el conocimiento del mercado, el conocimiento y normalización de los procesos de trabajo, la implantación de métodos objetivos de medición y la mejora de la gestión interna de las organizaciones⁴.

El concepto del control de la calidad está directamente asociado a la satisfacción de los usuarios. Existe una preocupación creciente en usuarios, productores y distribuidores sobre los servicios ofertados. Y, aunque a cada una de estas partes les mueven motivos diferentes, lo cierto es que las tres convergen en un mismo fin: acceder o proporcionar información precisa y exhaustiva sobre una materia, así como acceder o proporcionar servicios que agilicen el acceso a dicha información. En este último sentido, la preocupación en España alcanza al nivel institucional, pues como dice expresamente el III Plan Nacional de I+D (de aplicación entre 1996 y 1999) en su introducción:

Los países han de desarrollar políticas educativas, científicas, tecnológicas de comunicaciones, fiscales, etc., que sean coherentes y favorezcan el proceso de innovación y de difusión de tecnología y conocimientos, teniendo en cuenta sus características específicas.

Esta preocupación ha ido generando cada vez mayor número de trabajos cuya finalidad es analizar el grado de satisfacción/insatisfacción de los

² BROCKMAN, J. Just another management fad? The implications of TQM for library and information services. *ASLIB Proceedings*, 1992, vol. 44, n 7/8, p. 283-288.

³ Entre ellos, dos de los más representativos son SCOUG (Seven Continents Online User Group, antiguo Southern California Online User Group) y el Groupe Qualité des Banques de Données de l'ADBS.

⁴ GARCÍA-MORALES HUIDOBRO, E. Aportaciones de la gestión de calidad a bibliotecas y servicios de documentación. *Revista Española de Documentación Científica*, 1995, vol. 18, n^o 1, p. 1.

consumidores de información⁵. Muchos de los informes que se han publicado al respecto arrojan cifras no siempre halagueñas. En un informe del año 1989 sobre las bases de datos de uso profesional, la ADBS francesa constató una tasa de mortalidad de dichas bases de datos del orden del 16%. Constató, igualmente, que una gran parte de ellas no alcanzaba un índice de satisfacción adecuado entre sus usuarios. Por otra parte, una encuesta realizada por el IMO (Information Market Observatory) en 1991, mostraba que el 26% de los encuestados había anulado su suscripción. Entre las razones alegadas destacan: mala calidad del servicio, mejores prestaciones en la competencia, baja pertinencia en las búsquedas, cuestiones de índole presupuestaria, o uso insuficiente de las mismas. En otro informe realizado por este mismo organismo en el año 1995, referido a productos y servicios de información electrónica, se enfatiza claramente el hecho de que los usuarios de estos servicios demandan cada vez más la mejora en la calidad de estos productos⁶.

La razón por la que cada vez es mayor la importancia de los estudios sobre la calidad de la información reside en el incremento exponencial de su consumo. De hecho, esta preocupación no es exclusiva de las bibliotecas o los centros de documentación. El sector económico y comercial implicado en la producción y distribución de estos servicios se muestra cada vez más interesado en ello, por claras razones de mercado.

4. LAS BASES DE DATOS DE CIENCIAS SOCIALES

Hemos comentado ya que la gestión de la calidad está íntimamente relacionada con el contexto de actuación. En el caso concreto de las bases de datos, las materias que éstas abarcan son un factor determinante. Por una parte, las necesidades y expectativas de sus usuarios varían en función de éstas. Por otra, las características intrínsecas de cada disciplina condicionan el desarrollo de las fuentes utilizadas en el acceso y adquisición de conocimiento, así como el fondo y la forma de los documentos fruto de la investigación, lo que también incide directamente en el tratamiento y la gestión de estas fuentes.

La subjetividad que preside gran parte de los trabajos del ámbito de las Ciencias Sociales está en íntima relación con su objeto de estudio, que es el

⁵ DUFLOS, A. La qualité: enquête chez les producteurs de services électroniques d'information. 3^e Forum La qualité des services d'information documentaire. Paris, diciembre 1994, n 25.

LESTER, D. E. *The impact of quality management on the information sector: a study of case histories*. EUSIDIC, 1994.

AMSTRONG, C. J.; MEDAWAR, K. *Investigation into the quality of databases in general use in the U.K.* British Library Project, cotober 1994-april 1995.

⁶ INFORMATION Market Observatory (IMO). *The quality of electronic information products and services*. Luxembourg, September 1995. IMO Working Paper 95/4, p. 1-16.

hombre mismo en todas sus manifestaciones históricas, políticas, sociales, culturales y económicas. No cabe suponer que ninguna filosofía social pueda deslindarse de las relaciones que existen entre los problemas, las valoraciones, los hábitos o incluso los prejuicios de su época. El erudito en estas materias no puede hacer profesión de imparcialidad más allá de la fidelidad de las fuentes. La variedad posible de enfoques sobre un mismo tema dificulta la tarea de recopilación de los datos, y el elevado grado de dispersión de las fuentes dificulta el acceso al documento primario.

Es también una peculiaridad significativa de las Ciencias Sociales la gran fragmentación que existe entre ellas, que ha dado lugar a la aparición de muchas disciplinas y subdisciplinas. En ocasiones es difícil delimitar con claridad las diferentes áreas de estudio, puesto que en torno al hombre surgen muchas variables. Se trata, además, de disciplinas muy dinámicas, puesto que al ser el hombre el objeto de estudio, éste se ve sometido a un proceso de cambio y evolución constantes.

Exceptuando la Economía, la investigación en Ciencias Sociales no tiene un nivel alto de obsolescencia y sus estudios mantienen su vigencia durante largos periodos de tiempo. Esto supone que el volumen de información de interés se ve sometido a un proceso de acumulación que afecta al tratamiento y recuperación de la misma. Esta situación es la opuesta a la que nos encontramos en la información relativa a la tecnología, o a las Ciencias Experimentales y de la Salud, por ejemplo.

El interés que suscitan estudios que exigen un tratamiento extenso, unido a que la vida media de la información en esta área es más larga, hacen de las monografías una fuente de adquisición de conocimiento más valiosa que en otras disciplinas. Sin embargo, con los artículos de revistas la situación es la inversa, ya que su utilización es menor con respecto a otras ciencias. Así, mientras que en el campo de las experimentales el primer boletín de resúmenes data de 1830 —el *Pharmaceutisches Zentral Blatt*—, en 1898 ya comienza a editarse el *Physic Abstract*, en 1907 el *Chemical Abstracts* o en 1926 el *Biological Abstracts.*, hasta 1952 no aparece el primer boletín de resúmenes en Humanidades (cuyas características, en este sentido, son similares a las de las Ciencias Sociales), el *Historical Abstracts*, y es en 1965 cuando ve la luz el *Sociological Abstracts*. A este respecto, hay que tener en cuenta también que las disciplinas del área de las Ciencias Sociales gozan una institucionalización académica posterior. Asimismo, estudios llevados a cabo en el Reino Unido, constatan una proporción de citas a artículos de revistas frente a monografías muy inferior en las Ciencias Sociales con respecto a otras disciplinas.

Por otra parte, la literatura gris es un vehículo informal de comunicación muy importante en las Ciencias Sociales⁷. Ésta constituye un canal de

⁷ Artus menciona en su libro que, de acuerdo con una encuesta celebrada en la década de los ochenta en la entonces República Federal Alemana, un 50'1% de los especialis-

comunicación de resultados rápido y barato que no exige la sanción oficial favorable de los considerados expertos. Por último, en lo que se refiere a tipos de documentos, cabe señalar que la publicación en obras colectivas es frecuente en este tipo de ciencias.

La terminología empleada en estas áreas del conocimiento es ambigua, imprecisa, provisional e inestable y carece de la exactitud que rige la de otras ciencias. El significado de los términos cobra acepciones diferentes según la materia, la época o el contexto en el que se utiliza. Incluso, dentro de una misma materia, las diferentes escuelas utilizan los vocablos con diferentes matices. El uso del lenguaje cotidiano con significados especiales y específicos también es característico de los trabajos de Ciencias Sociales, así como la variedad de significados según los países y culturas. La inestabilidad y la falta de normalización en la terminología plantean graves problemas tanto en el desarrollo de tesauros como en la indización de la información⁸.

Otra característica es la falta de consenso respecto a los métodos de recogida, análisis e interpretación de los datos, así como respecto a los problemas y la forma de tratarlos. Es un hecho común que las Ciencias Sociales cuentan con cientos de teorías nunca suficientemente demostradas y nunca suficientemente rechazadas, lo que plantea un enorme problema a la hora de proporcionar información pertinente.

Otra característica de la información en Ciencias Sociales es su escasa rentabilidad a corto plazo, que conlleva una perenne falta de recursos para llevar a cabo proyectos de sistematización de la información. Esto repercute de forma directa en el desarrollo de las fuentes, tanto impresas como electrónicas.

En la actualidad, no es difícil demostrar que la información y el conocimiento de las Ciencias Sociales no superan fácilmente las fronteras sociales, culturales, lingüísticas, políticas o económicas. Los datos pueden transferirse a cualquier parte del mundo con una facilidad cada vez mayor. Sin embargo, en este caso, su pertinencia y aplicabilidad no se pueden deslindar, en una gran mayoría de casos, del preciso contexto en el que, o para el que, se han llevado a cabo los estudios. Las Ciencias Sociales, desde un punto de vista general, carecen del carácter universal del que gozan otras ciencias, y esto hace que la información sea poco recogida en otros países distintos al productor de la información.

tas en Ciencias Sociales pertenecientes a Centros de Investigación de este país estimó que este tipo de información constituye la vía más importante para la difusión de los resultados de sus trabajos. ARTUS, M. *Grave Literatur als Medium Wissenschaftlicher Kommunikation*, 1984.

⁸ SAHLI, M. S. *Terminology of the Social Sciences: the Term Cognitive Processes in the Thesauri of Two Discipline-Based Information Systems*. Michigan: UMI, 1981.

RIGGS, F. W. *A New Paradigm for Social Science Terminology*. UNESCO, INTERCONCEPT Project, Noviembre 1979, vol. 6, n° 3, p. 150-158.

Es también necesario, a la hora de estudiar las bases de datos de estas ciencias, considerar el problema de la acumulación anárquica de datos, documentos y conocimientos. Muchos científicos sociales siguen una vía personal en su investigación y ello dificulta enormemente la recopilación de información.

En cuanto a la tipología del usuario de información en esta área del conocimiento, cabe señalar que aunque el investigador demande un tipo de información muy concreta y especializada, normalmente sus necesidades informativas abarcan un gran número de campos de interés. Por otra parte, Michael Brittain observó que los investigadores y estudiosos de las Ciencias Sociales, al igual que los de Humanidades, se valían, principalmente, de fuentes impresas en sus búsquedas de información⁹. Esto, sin duda, está relacionado con el hecho de que las bases de datos de estas disciplinas, salvo en las de Economía, han presentado menos ventajas y más inconvenientes que en las de ciencia y tecnología. Al utilizarse menos, la rentabilidad económica es menor, razón por la cual se invierte menos en mejorarlas, y, por tanto, su utilidad es también menor.

Sin embargo, cabe señalar que este panorama se está viendo alterado últimamente, puesto que se observan cambios significativos en los usos de la información de los científicos de esta materia. Así, por ejemplo, en el año 1995, las bases de datos en línea de Ciencias Sociales del Consejo Superior de Investigaciones Científicas se utilizaron dos veces más que las de Ciencias Experimentales, Tecnología o Ciencias de la Salud. Y aunque esto se debe, entre otras razones, a que los investigadores de estas otras áreas, recurren más a fuentes extranjeras, sí constata el hecho de que cada vez son más los estudiosos de las Ciencias Sociales los que hacen uso de recursos electrónicos.

Además, los intentos de los documentalistas y bibliotecarios por clasificar los contenidos de las distintas disciplinas y subdisciplinas de las Ciencias Sociales, así como para normalizar su terminología, están dando frutos interesantes. Se cuenta ya con clasificaciones *ad hoc* y con un amplio repertorio de Tesoros, que sin duda repercuten en una mejora de la calidad de estas bases de datos que, a su vez, implican un mayor uso de las mismas.

5. INDICADORES DE LA CALIDAD

Existen toda una serie de indicadores que sirven para determinar la calidad de una base de datos documental. Estos indicadores pueden clasificarse en dos grandes grupos: aquéllos que se refieren a los aspectos formales de

⁹ BRITTAİN, M. Fronteras culturales de las ciencias sociales en los años noventa: nuevas políticas de documentación, información y creación de documentos. *Revista Internacional de Ciencias Sociales*, vol. 119, p. 111-120.

los registros y los que se centran en cuestiones que afectan al contenido. Durante las dos últimas décadas, los trabajos sobre control de calidad de bases de datos se han orientado más hacia el primer tipo de indicadores. Sin embargo, el análisis de las cuestiones relacionadas con el contenido es prioritario en este tipo de bases de datos.

Además de estas dos clases de indicadores, el control de calidad de una base de datos documental tiene que tener en cuenta el software de recuperación, la interfaz de usuario, así como los servicios ofertados por los productores y/o distribuidores.

5.1. *Indicadores formales*

Los indicadores formales utilizados para evaluar la calidad de una base de datos se refieren, principalmente, a la tasa de errores contenidos en los campos de los registros.

La detección de errores ha sido objeto de estudio desde hace tres décadas, durante las cuales se ha analizado el lenguaje con la ayuda de los ordenadores, lo que ha dado lugar, además, a la aparición de la lingüística computacional como nueva disciplina. De hecho, existen test de calidad que se centran, exclusivamente en el análisis del índice de errores.

Los llamados errores pueden ser de diversas clases. Jacsó los clasifica en los que él denomina errores de omisión y de «comisión»¹⁰. Los primeros se refieren a la ausencia de datos y/o de campos. Los segundos, a incorrecciones tipográficas u ortográficas. Existen, además, errores de tipo gramatical y semántico, que no suelen estimarse debido a su baja repercusión en la recuperación.

Entre los errores tipográficos pueden definirse seis clases: errores por omisión, inserción, sustitución, permutación, repetición y espacios en blanco. Son muchos los autores que han estudiado su frecuencia de aparición¹¹, y todos ellos coinciden en afirmar que su orden de aparición es el siguiente:

¹⁰ JACSÓ, P. Searching for skeletons in the database cupboard. Part I: errors of omission. *Database*, 1993, vol. 16, n 1, p. 67-76.

JACSÓ, P. Searching for skeletons in the database cupboard. Part II: errors of commission. *Database*, 1993, vol. 16, n 2, p. 77-82.

¹¹ Entre otros:

O'NEILL, E.T.; VIZINE-GOETZ, D. The impact of spelling errors on databases and indexes. *Proceedings of the National Online Meeting*, New York, 1987, May 9-11, p. 313-320.

O'NEILL, E.T.; VIZINE-GOETZ, D. Quality control in online databases. *Annual Review of Information Science and Technology*, 1988, vol. 23, p. 125-156.

PULLOCK, J. J.; ZAMORA, A. Collection and characterization of spelling errors in scientific and scholarly text. *Journal of the American Society for Information Science*, 1980, vol. 34, n 1, p. 51-58.

ZAMORA, A. Automatic detection and correction of spelling errors in a large database. *Journal of the American Society for Information Science*, 1980, vol. 30, n 1, p. 51-57.

te: entre un 30-40% son de omisión, entre un 25-30% de inserción, entre un 15-25% de sustitución, entre un 10-15% de permutación, entre un 20-25% de repetición, y entre un 25-30% son espacios en blanco. Los expertos en esta materia tienden a considerar como aceptable un porcentaje del 0,3% del total de las palabras como errores admisibles¹².

Para llevar a cabo un análisis que permita determinar el índice de errores se trabaja con los campos textuales de una muestra representativa de registros. De las distintas metodologías que existen para analizar la tasa de error, la propuesta de Bourne¹³ se ha utilizado en muchos de los trabajos sobre este tema¹⁴.

5.2. Indicadores de contenido

El control de calidad referido al contenido de una base de datos documental se centra en distintos aspectos. Se refiere, por un lado, al análisis de la cobertura temática, cronológica, geográfica y de tipología documental. Se ocupa, igualmente, de la estructura y contenido de los registros bibliográficos, así como de la existencia de posibles duplicados. El análisis de la autoría de los trabajos a los que se hace referencia es un aspecto fundamental que también tiene que tenerse en cuenta, así como el de la indización de los documentos.

5.2.1. Análisis de la cobertura

El análisis de la cobertura es un proceso complejo, ya que, por un lado, está directamente relacionado con el propósito de la misma. Esto es, con la finalidad para la que ha sido diseñada, y con el tipo de usuarios a los que está destinada; factores ambos que van a condicionar la evaluación de los distintos parámetros. Por otro, porque se trata de un proceso donde la relación existente entre cada una de las variables es lo que va a determinar el análisis final.

En principio, el tamaño de la base de datos puede servir como un indicador aproximativo para estimar la adecuación de su cobertura temática, en función del periodo cronológico que abarca. Asimismo, el grado de genera-

¹² SPINAK, E. Errores ortográficos en el ingreso en bases de datos. *Revista Española de Documentación Científica*, 1995, vol. 18, n 3, p. 307-319.

¹³ BOURNE, E. C. Frequency and impact of spelling errors in bibliographic databases. *Information Processing & Management*, 1977, vol. 13, n 1, p. 1-12.

¹⁴ Un ejemplo de aplicación de esta metodología es el trabajo llevado a cabo por HERRERO SOLANA, V. La calidad total en bases de datos españolas: estudio de la tasa de error en las bases del CSIC. *Revista Española de Documentación Científica*, 1997, vol. 20, n 4, p. 409-416.

lidad o especificidad de la materia también va a influir en la perspectiva del análisis. La cobertura temporal, así como la actualidad de los datos, tienen que estar en consonancia con la temática de la base de datos. En el caso de las Ciencias Sociales, las búsquedas retrospectivas son más importantes que en otras disciplinas. En este punto, cabe hacer una salvedad, y es el caso de la Economía, en la cual la obsolescencia de la información es mayor. Así, por ejemplo, en las Ciencias Políticas el índice de obsolescencia se sitúa entre tres y siete años, mientras que en Economía está entre dos y cinco.

A principios de los años sesenta comienza a utilizarse como índice del envejecimiento de la información científica el concepto de semiperiodo, definido por Burton y Kebbler como el «tiempo durante el cual se ha publicado la mitad de la información científica que se utiliza en un momento dado y en determinada especialidad»¹⁵. Es un concepto imprescindible para obtener una medida precisa del envejecimiento de la información. Su determinación se basa, igualmente, en el análisis de las referencias bibliográficas distribuidas por años de procedencia, considerando las publicaciones citadas en un momento dado como la información activa circulante en el mismo. El semiperiodo, así considerado, coincidirá con el tiempo estimado que tardará en dejar de utilizarse el 50% de la información publicada en un momento dado.

Los valores del semiperiodo varían considerablemente de unas ciencias a otras. Hay disciplinas científicas que envejecen con mucha rapidez (la Física o la Ingeniería, por ejemplo), con semiperiodos del orden de tres a cinco años. Otras, como las Matemáticas o la Botánica, presentan semiperiodos más largos, en torno a los diez años. En líneas generales, la información en Ciencias Sociales tiene un periodo de interés que oscila entre valores de tres a siete años. Sin embargo, como explica Álvarez-Ossorio, con el aumento de trabajos sobre envejecimiento de la información científica, se han comenzado a observar resultados dispares, lo que ha servido para concluir que el semiperiodo debe considerarse un concepto relativo y no absoluto, como se postulaba en un principio¹⁶.

En cuanto a la cobertura geográfica de la base de datos, cabe señalar que en el caso de las Ciencias Sociales, ésta suele ceñirse a estudios de ámbito más restringido; sus usuarios requieren trabajos de áreas geográficas concretas— es obvio que cuestiones políticas o sociológicas, por ejemplo, difíciles de extrapolar a diferentes países—. Este hecho supone que países distintos al productor de la información muestren un escaso interés por ello. Es usual que la universalidad no presida la cobertura geográfica de estas bases de datos, pero, en principio, no tiene porqué ser considerado

¹⁵ BURTON, R. E.; KEBBLER, R. W. The half-life of some scientific and technical literatures. *American Documentation*, 1960, vol. 11, p. 18-22.

¹⁶ PÉREZ ÁLVAREZ-OSSORIO, J. R. *Introducción a la Información y Documentación Científica*. Madrid: Alhambra, 1990, p. 17-19.

un indicador negativo. En el caso de las Ciencias Sociales, son más valiosas que en otras disciplinas las bases de datos nacionales, ya que, de lo contrario, es difícil acceder a información de interés para los investigadores de esta materia.

Esto va a condicionar, igualmente, la procedencia geográfica de las revistas recogidas en estas bases de datos: una gran parte de ellas suelen estar editadas en el país productor de la información. Existe la idea generalizada de considerar de baja calidad las publicaciones nacionales de los países no anglosajones, dado que muchas de ellas no se incluyen en índices internacionales. En el caso de estas ciencias, la publicación en revistas de ámbito nacional es mucho más habitual, por lo que la causa de que no se incluyan en índices internacionales no puede considerarse, a *priori*, un indicador de baja calidad.

No obstante, para evaluar la calidad de las revistas de las que se nutren las bases de datos es habitual recurrir a estos índices. El más utilizado es el *Journal Citation Report*, editado por el Institute of Scientific Information (ISI) de Filadelfia, que se sirve de técnicas bibliométricas para reflejar la proyección internacional de las publicaciones. Para ello utiliza el número total de citas producidas y el factor de impacto. Hay que tener en cuenta, sin embargo, que, por un lado, las revistas que publican menor número de artículos pero de gran extensión se ven favorecidas por este indicador, y, por otro, que éste presenta variaciones significativas según la disciplina de la que se trate, el tamaño de la Comunidad Científica, los ámbitos y hábitos de citación o el envejecimiento de la información. En general, las publicaciones de investigaciones básicas y de colectivos muy numerosos son las que presentan mayor factor de impacto.

En lo que respecta a la tipología documental recogida en estas bases de datos, no pueden utilizarse los mismos parámetros válidos para otras disciplinas. Como ya hemos comentado, en estas ciencias las monografías y la literatura gris cobran mayor relevancia. Es importante, pues, que este tipo de publicaciones estén recogidas en la base de datos en cuestión. Con el fin de poder evaluar este parámetro es conveniente que los registros contengan el campo «tipo de documento», información, además, de gran relevancia a la hora de delimitar las búsquedas.

5.2.2. *Análisis de los registros*

No cabe duda de que la cantidad y el tipo de información que proporcionan los registros bibliográficos son un claro indicador de la calidad de una base de datos. La exhaustividad de la información es, en principio, un aspecto deseado. Sin embargo, esta exhaustividad no debe repercutir en una redundancia de información. La existencia de campos de información repetitiva o innecesarios inciden negativamente en la evaluación de la cali-

dad de una base de datos, ya que éstas tienen que atenerse al principio de brevedad que debe regir la presentación de información en forma electrónica.

Debe existir, igualmente, un equilibrio entre el número y tipo de campos de los registros, ya que un exceso de campos de información meramente descriptiva puede derivar en resultados poco pertinentes en la búsqueda.

Ahora bien, las bases de datos documentales tienen que satisfacer necesidades de información de muy diversa índole: desde una simple consulta de referencia hasta búsquedas exhaustivas de orientación bibliográfica. Los campos descriptivos proporcionan información al margen del contenido de los documentos y sirven, principalmente, para delimitar las búsquedas, para resolver demandas puntuales, así como para llevar a cabo estudios bibliométricos o estadísticos.

En el caso de las Ciencias Sociales, donde predominan las búsquedas temáticas, los campos que proporcionan información de contenido son más relevantes, por lo que, además del campo de título, los de valor añadido, tales como el de descriptores, resumen o notas, son los más utilizados. En la evaluación de la calidad de estas bases de datos, el análisis de estos campos es imprescindible. Así, por ejemplo, los registros deben mostrar el título completo de los documentos, ya que se trata de un campo de gran valor para la recuperación de información por lenguaje libre. Los campos de valor añadido proporcionan puntos de acceso fundamentales en la recuperación de información. El resumen, además, permite a usuario discernir sobre el hipotético interés del documento. La uniformidad y consistencia en su tratamiento son también indicadores de la calidad de la base de datos.

5.2.3. *Detección de duplicados*

La duplicidad de los registros presenta dos formas distintas: registros idénticos referidos a un mismo documento y registros diferentes referidos a un mismo documento. Los primeros suelen ser escasos, ya que cada registro posee tal número de campos y de información que es difícil que se introduzcan más de uno de forma idéntica. Además, son fáciles de detectar por los sistemas de gestión de bases de datos¹⁷. El segundo caso es el que más veces se presenta y el más difícil de paliar.

La duplicidad de registros aumenta innecesariamente el volumen de la base de datos, y dificulta el acceso a la información, ya que el usuario se ve obligado a seleccionar los más idóneos entre un volumen más grande. Un índice alto de duplicados es un claro indicador negativo, por lo que es fundamental que su detección se lleve a cabo durante el proceso de alimenta-

¹⁷ O'NEILL, E. T.; ROGERS, S.; OSKIN, W. M. Characteristics of duplicate records in OCLCs Online Union Catalogue. *Library Resources and Technical Services*, 1993, vol. 37, n 1, p. 60.

ción de la base de datos mediante técnicas de validación. Su detección a posteriori es complicada y costosa. De hecho, grandes bases de datos del tipo OCLC, han optado por no llevar a cabo esta tarea íntegramente, ya que no les resulta rentable.

En el proceso de detección de duplicados se trabaja, predominantemente, con campos numéricos y específicos de cada documento, tales como el ISBN o ISSN¹⁸.

5.2.4. *Análisis de la autoría*

En principio, el prestigio científico, la proyección internacional y la procedencia institucional de los autores de los documentos a los que se refiere una base de datos es un indicador de la calidad de la misma. Sin embargo, se trata de parámetros difíciles de evaluar, ya que aunque los instrumentos que sirven para medirlos se basan en criterios objetivos, existen variables de muy diversa índole que influyen en los resultados.

Una forma de «cuantificar» el prestigio y la proyección de los autores consiste en recurrir a bases de datos de citas. En este caso, el *Social Citation Index*, del ISI, que proporciona un índice de autores, con sus correspondientes trabajos, bajo cada uno de los cuales aparece el conjunto de artículos en los que han sido citados. Se completa con un índice de materias permutado y un índice de fuentes donde aparecen los trabajos citantes, ordenados por autores e instituciones donde se realizaron. En el caso de esta base de datos, sólo se recogen las citas del primer firmante, por lo que el resto se ve perjudicado.

Los análisis de citas de los investigadores o grupos de investigación resultan de un interés indudable no sólo para evaluar la calidad de las bases de datos, sino también, por ejemplo, para planificar la política científica de un país. Ahora bien, los recuentos de citas deben interpretarse con cautela. Las razones por las que un autor es más o menos citado no siempre guardan una relación directa con la calidad de sus investigaciones.

El trabajo en equipo favorece la proliferación de citas, ya que los colegas tienden a citarse entre ellos. En las Ciencias Sociales predomina, frente a otras disciplinas, el trabajo en solitario. Las revistas de ámbito anglosajón son más consultadas por la comunidad científica, por lo que los que publican en ellas tienen más probabilidades de ser citados. Hemos hablado ya del carácter local de las investigaciones en Ciencias Sociales, que no favorece la publicación en revistas extranjeras. Sabido es, además, que Ciencia, con mayúsculas, y ciencia institucional, muchas veces no siguen caminos paralelos. Quienes no pertenecen a grupos de investigación institucional-

¹⁸ TONEY, V. Cleanup and the duplication of an International Bibliographic Database. *Information Technology and Libraries*, 1992, vol. 11, p. 21-23.

mente respaldados tienen, también, menos posibilidades de acceder a los canales regulares de difusión de la información.

Es, por tanto, importante, efectuar este tipo de análisis estrictamente dentro de una misma disciplina, pues las pautas de citación varían sensiblemente. Así, en las bases de datos multidisciplinares que aunque se corresponden con una sólo área de la ciencia abarcan diferentes disciplinas, frecuentes en las Ciencias Sociales, la muestra debe ser representativa de todas ellas.

En cuanto a la productividad de los autores, ésta también varía según la disciplina en la que trabajen. En las Ciencias Sociales, una categorización aproximada sería la de considerar como «grandes productores» a quienes han publicado diez o más trabajos, «medianos productores» los que cuentan con entre dos y diez, y «pequeños productores» los que tienen una o dos publicaciones. Este último grupo es el que suele engrosar la mayoría de las bases de datos en esta área del conocimiento, por lo que la evaluación de la calidad ha de centrarse en los dos primeros grupos.

Se debe advertir, también, que la productividad de los científicos no tiene por qué, necesariamente, coincidir con la calidad de sus trabajos, si bien hay muchos, como Price¹⁹, que sí están de acuerdo entre la correlación de ambos parámetros. Quizás, más que de la productividad, habría que hablar del impacto que sus trabajos han producido sobre autores posteriores, para lo cual es necesario recurrir a estudios bibliométricos de citas, proceso complicado, que consume gran cantidad de tiempo, por lo que no suele llevarse a cabo en la evaluación de una fuente electrónica.

La procedencia institucional de los autores recogidos en una base de datos es, también, un indicador de la calidad de la misma. Por ello, es interesante que los registros contengan el campo de lugar de trabajo, muchas veces ausente en las bases de datos de Ciencias Sociales.

Dijimos ya que las obras colectivas constituyen una forma de comunicación más frecuente en las ciencias sociales que en otras disciplinas. Cuando éste es el caso, algunas bases de datos dejan el campo de autor vacío (quizás influenciadas por criterios catalográficos), hecho que incide negativamente en la recuperación documental cuando se efectúan búsquedas por autor.

5.2.5. *Análisis de la indización*

Si por indización entendemos la operación que consiste en escoger los términos más apropiados para representar el contenido de un documento, la pertinencia e idoneidad de los mismos tienen que considerarse claves en la evaluación de la calidad de una base de datos. Es cierto que las capacida-

¹⁹ PRICE, D. J. S. Networks of scientific papers. *Science*, 1965, n 149, p. 510-515.

des de los softwares para recuperar por lenguaje libre han paliado los problemas derivados de indizaciones inapropiadas desde distintos puntos de vista, ya que posibilitan el acceso a campos como el de título o resumen. Sin embargo, también es cierto que el campo de materias es una de las fuentes más valiosas en las búsquedas documentales. La calidad de la indicación es, pues, fundamental.

En nuestra opinión, la primera cuestión que debe tenerse en cuenta a la hora de evaluar la calidad de la indización es que la base de datos utilice descriptores y no encabezamientos de materia. La ineficacia de estos últimos en entornos automatizados está sobradamente demostrada. Respecto a los descriptores empleados, hay que analizar si respetan o no las normas de construcción de los mismos.

La evaluación de la calidad de la indización exige elegir muestras representativas, establecer un nivel aceptable de fallos, examinar las muestras y decidir la aprobación o el rechazo al nivel de calidad aceptable.

El análisis de este parámetro se rige por tres principios fundamentales: consistencia, relevancia y exhaustividad²⁰. El principio de consistencia establece que cada concepto debe expresarse siempre con el mismo descriptor e idéntica morfología. La relación entre concepto y descriptor ha de ser biunívoca: a cada concepto un descriptor y a cada descriptor un concepto. El principio de relevancia mide la exactitud con la que un concepto está representado por un término de indización. La exhaustividad está relacionada con el número de nociones que caracteriza el contenido íntegro del documento y el número de descriptores empleados para describir esos conceptos.

El análisis de la consistencia se realiza mediante la elección de grupos o racimos —*clusters* en su terminología inglesa— de documentos de contenido temático similar, para lo cual se recurre a los campos de título y resumen, nunca a los descriptores, ya que son éstos el objeto de la evaluación. Conviene que la mitad de los racimos abarque temas generales y la otra mitad específicos. Una vez seleccionados los racimos, se identifican los descriptores presentes en más de un registro, con el fin de seleccionar los que se repiten en la mitad o más de los registros de cada racimo. Estos últimos serán los que se atengan al principio de consistencia²¹.

Para conocer el grado de relevancia de un término se utilizan los llamados «valores de discriminación» que se calculan dividiendo el número de registros que contienen un descriptor determinado entre el total de registros de la base de datos²². Se estima que un valor aceptable de discrimina-

²⁰ WULFF BARREIRO, E. Calidad de la indización en bases de datos de tamaño cualesquiera. *Boletín de la Asociación Andaluza de Bibliotecarios*, 1993, vol. 9, n 31, p. 41-45.

²¹ WHITE, H. D.; GRIFITH, B. C. Quality of indexing in online databases. *Information Processing and Management*, 1987, vol. 23, p. 211-214.

²² CHU, C. M.; ALJIFERUQUE, I. Quality of indexing library and information science databases. *Online Review*, 1989, vol. 13, n 1, p. 11-35.

ción ronda el 0.05, ya que se corresponde con más del 5% de los documentos de cualquier base de datos. Para calcular la relevancia de la indización no es conveniente utilizar descriptores muy específicos que aparecen en pocos registros, ni tampoco los muy generales que se encuentran en muchos.

Por último, la evaluación de la exhaustividad se calcula obteniendo la media del número de descriptores utilizados. Para ello, basta con dividir el total de descriptores de la muestra entre el número de registros de ésta. Se considera un nivel aceptable de exhaustividad entre ocho y doce descriptores por documento.

6. SOFTWARE DE RECUPERACIÓN E INTERFAZ DE USUARIO

Hemos dicho ya que consideramos una base de datos como un producto concreto de la industria de la información, producto que incluye, además de la propia información, la forma en la cual ésta puede ser recuperada. La evaluación de una base de datos puede ceñirse, únicamente, a la información que contiene, sin embargo, la del producto final tiene, necesariamente, que incluir el análisis del software de recuperación y el de la interfaz de usuarios. Del primero dependen las capacidades de búsqueda del sistema y de manipulación de los resultados. De la interfaz, la facilidad de uso y comprensión de las instrucciones, menús o formularios que permiten efectuar las búsquedas. Desde nuestro punto de vista, la interfaz constituye un aspecto fundamental, ya que la tendencia actual es que los recursos informativos puedan ser utilizados, directamente, por usuarios finales.

6.1. *Software de recuperación*

En la evaluación del software es útil recurrir a formularios predefinidos. La bibliografía al respecto recoge un gran número de ellos²³. Se trata, en

²³ Entre otros, caben señalar los siguientes:

HARRY, V.; OPPENHEIM, Ch. Evaluations of electronic databases. Part I: Criteria for testing CD-ROM products. *Online & CD-ROM Review*, 1993, vol. 17, n 4, p. 211-222.

HARRY, V.; OPPENHEIM, Ch. Evaluations of electronic databases. Part II: Testing CD-ROM products. *Online & CD-ROM Review*, 1993, vol. 17, n 6, p. 339-347.

SCHWARTZ, C. Evaluating CD-ROM productsÑ Yet another list. *CD-ROM Professional*, 1993, vol. 4, n 1, p. 87-105.

SIEVERTS, E. G.; HOFSTEDE, M. Software for information storage and retrieval tested, evaluated and compared. Part I: General introduction. *The Electronic Library*, 1991, vol. 9, n 3, p. 145-153.

VÁLAS, G. Comparison of some widespread CD-ROM information retrieval packages. *Online & CD-ROM Review*, 1994, vol. 18, n 4, p. 211-226.

general, de relaciones muy similares en el contenido, si bien es posible encontrar ligeras variaciones en su forma y organización. Los aspectos fundamentales susceptibles de evaluarse se refieren a las capacidades de búsqueda, a los formatos de salida de datos y a la fiabilidad y seguridad del sistema.

En cuanto a lo primero, se espera que los sistemas de recuperación de bases de datos documentales hagan posible la manipulación de conjuntos de resultados, así como el uso de operadores booleanos, de adyacencia y de proximidad, y del truncamiento en las estrategias de búsqueda. Se requiere, igualmente, que posibiliten la recuperación por lenguaje libre y controlado, así como el acceso en línea a los índices y la selección directa de sus componentes. El proceso de búsqueda debe permitir guardar y reutilizar tanto estrategias como perfiles de búsqueda concretos.

En lo que respecta a la salida y manipulación de datos, es importante que el sistema proporcione distintos tipos de formatos, y posibilite al usuario definir los suyos propios. Se requiere que permitan importar resultados e históricos de búsqueda, para que puedan ser manipulados posteriormente por cualquier procesador de textos. Es imprescindible que el usuario pueda seleccionar referencias en función de diferentes parámetros.

Y en cuanto a la fiabilidad y seguridad del sistema, se concede especial importancia a cuestiones tales como la aparición de mensajes de error, a las interrupciones en las búsquedas y a la facilidad para reconectar el software.

6.2. *Interfaz de usuario*

El término interfaz hombre-máquina (*human-computer interface*, en su terminología anglosajona) podría definirse como el conjunto formado por las palabras y los símbolos que se ven en las pantallas del ordenador; el contenido y los signos ortográficos de los formatos de presentación, los procedimientos o mecanismos utilizados para ejecutar, almacenar y obtener información; y la estructura de organización de la interfaz como un todo²⁴.

La interfaz especifica la forma en la que el programa y el usuario se comunican, por lo que los objetos de estudio y evaluación de una interfaz son: la estructura y forma de los diálogos de transacción, las pantallas a las que el usuario tiene acceso, la presentación de los resultados de las búsquedas y de los registros bibliográficos, la ayuda en línea y los mensajes que el sistema emite al usuario. La presentación y disposición de todos ellos forma un conjunto que determina la facilidad y amigabilidad del sistema y, por tanto, las posibilidades apriorísticas de éxito/fracaso que puede esperar el usuario de dicho sistema.

²⁴ DUMAS, J. *Designing user interfaces for software*. Englewood Cliffs, NJ: Prentice-hall, 1988.

La consistencia y la claridad son los dos criterios que deben presidir el diseño de una interfaz de usuario en su globalidad. Debe prestarse especial atención al principio de consistencia, ya que la consistencia es el pilar de los sistemas fáciles de utilizar, y fáciles de memorizar y recordar. Dicho principio debe presidir todos los aspectos que entran en juego en el diseño de una interfaz de usuario no profesional. La claridad minimiza la necesidad de actuar de los usuarios; la consistencia le ayuda a recordar los pasos que debe seguir y las instrucciones que debe ejecutar.

Es igualmente preciso que el sistema sólo brinde al usuario la información necesaria y en la forma en la que ésta puede ser manipulada directamente por él.

Las pantallas constituyen la primera consideración que debe tenerse en cuenta a la hora de analizar y evaluar una interfaz. Su estructura, organización y diseño inciden directamente en el proceso de búsqueda de los usuarios. En principio, la estructura debe reunir todos aquellos aspectos que se consideran necesarios en un buen diseño visual. Así, debe estar compensada, ser regular, simétrica, económica y secuencial. Siguiendo el principio de economía y las preferencias de los usuarios se recomienda que sólo se utilice el 30% del espacio físico disponible²⁵. Sobre el número máximo de caracteres por línea las recomendaciones oscilan entre 46 y 60.

Deben seguirse criterios normalizados para la ubicación de la información, incidiendo en el principio de consistencia. El lugar elegido para cada elemento debe ser siempre el mismo, con el fin de no despistar al usuario y facilitarle al máximo su localización. De la misma manera, la distinta tipología y las diferentes formas de resaltar los términos y opciones deben utilizarse de forma homogénea y consistente.

Las pantallas que muestran los resultados del proceso de búsqueda son de dos tipos: las que presentan los distintos índices y las de los registros bibliográficos. El propósito de las primeras es permitir a los usuarios acceder a los resultados de sus búsquedas y realizar selecciones rápidas de los términos de cada relación. El acceso a los registros bibliográficos debe ser posible en formatos breves y completos. Es recomendable que, de alguna manera, el sistema resalte el elemento de los registros que ha dado lugar al resultado correspondiente. Siempre que sea posible, debe limitarse la presentación del registro bibliográfico a una sólo pantalla.

En cuanto a los diálogos de transacción, éstos pueden consistir en un lenguaje de comandos, en menús, en formularios o en gráficos e iconos. La idoneidad de cada uno de ellos depende del tipo de usuarios. Los inexpertos suelen preferir los menús y formularios, frente a los expertos, que se inclinan por los comandos. Una combinación de ambos es, quizás, la opción más adecuada.

²⁵ GALITZ, W. *Handbook of Screen Format Design*, 3rd ed. Wellesley, MA: QED Information Sciences, 1989.

En cuanto a los mensajes que el usuario recibe del sistema, éstos son de distintos tipos, y todos ellos inciden directamente en el proceso de búsqueda. Los mensajes de respuesta reclaman al usuario que actúe. Si están bien diseñados reducen la probabilidad de error de los usuarios, ya que minimizan la cantidad de información que debe memorizar y le guían en el proceso de búsqueda. Los mensajes sobre el estado del proceso mantienen al usuario informado sobre el estado de la búsqueda. Son apropiados cuando los resultados tardan en aparecer, ya que reducen el estado de ansiedad propio de muchos usuarios, especialmente los inexpertos. Los mensajes de error alertan al usuario de que el sistema no es capaz de responder a su estrategia. La finalidad de estos mensajes no debe quedarse en advertirle del error, sino en guiarle sobre lo que puede hacer después.

La ayuda en línea es un componente esencial de las interfaces de usuario. La ayuda en línea no debe estar concebida como el equivalente electrónico a un manual impreso. Por el contrario, debe tratarse siempre de una ayuda contextualizada; esto es, responder a las peticiones del usuario según el momento preciso de su proceso de búsqueda.

7. SERVICIOS DE VALOR AÑADIDOS AL PRODUCTO

Por último, vamos a comentar someramente qué tipo de servicios proporcionados por los distribuidores o proveedores de información añaden calidad a una base de datos documental. De hecho, en muchas ocasiones, éste tipo de servicios llegan a determinar la selección de uno u otro producto.

En las bases de datos bibliográficas referenciales, es deseable que el distribuidor proporcione un servicio de acceso al documento primario, que debe aprovechar al máximo las nuevas tecnologías de la información con el fin de agilizar el proceso y abaratar los costes. La posibilidad de obtener traducciones supone, además, un servicio de gran valor.

El servicio de difusión selectiva de la información ha sido, tradicionalmente, uno de los más valorados, ya que permitía a los usuarios mantenerse actualizados sin necesidad de recurrir periódicamente a intermediarios. Sin embargo, en la actualidad, los usuarios están mucho más familiarizados con este tipo de fuentes electrónicas y los softwares, asimismo, han simplificado los procedimientos de búsqueda. Por ello, los usuarios tienden, cada vez con mayor frecuencia, a ser ellos mismos quienes llevan a cabo las consultas a las bases de datos de una forma periódica. Por ello, el valor del servicio de difusión selectiva de la información se está devaluando de forma progresiva, si bien sigue siendo apreciado por determinado tipo de usuarios.

Se valora también que el distribuidor/proveedor disponga de un servicio personalizado de asistencia al usuario, con interlocutores expertos en

distintas lenguas y materias. La asequibilidad y facilidad para la comunicación son fundamentales, por lo que el uso de mensajerías de correo electrónico es importante. Las cuestiones de asesoría y asistencia técnicas son especialmente valiosas, ya que pueden llegar a convertirse en problemas infranqueables para profesionales inexpertos en cuestiones informáticas.

Es importante, igualmente, que el distribuidor mantenga al día al usuario-cliente sobre las innovaciones o cambios, tanto del sistema de recuperación como del contenido de la base de datos.

Se considera fundamental que el usuario reciba, de forma gratuita y periódica, cursos de iniciación y formación continuada en el uso de la base de datos. Cuanto mayor sea la cobertura geográfica de estos cursos, más valor adquieren. No cabe duda que con ofrecerlos no basta, su calidad es lo más importante. Y en ella hay que valorar la exhaustividad y cobertura de la información que proporcionan, las distintas tipologías de usuarios a los que van dirigidos, los niveles de formación, el propio profesorado o la gratitud completa o parcial, por ejemplo.

Y un punto aparte merece el análisis del manual de usuario así como del resto de la documentación aportada por el distribuidor o proveedor en relación a los distintos elementos de la base de datos. Es importante evaluar su comprensibilidad y exhaustividad, el detalle y cobertura en la descripción de la base de datos y su cobertura, los ejemplos que proporciona, la adecuación y comprensibilidad de la terminología, la utilidad de los índices, así como su grado de actualización en función de las modificaciones del software y la base de datos.

8. REFLEXIÓN FINAL

A lo largo de este trabajo hemos expuesto las principales consideraciones que deben tenerse en cuenta a la hora de abordar el control de calidad en una base de datos de ciencias sociales. Como ha quedado patente, se trata de un proceso complejo, ya que, por un lado, son muchos los indicadores que deben analizarse. Por otro, entran en juego valores en íntima relación con las características específicas y propias de estas ciencias.

La complejidad de este proceso se acentúa por el hecho de que la adopción de criterios de índole meramente cuantitativa y/o extrapolados de otras áreas de conocimiento deriva en una distorsión de los datos. Se trata, además, de valores relativos, nunca absolutos, en cuyo análisis debe tenerse en cuenta que ningún indicador, por sí mismo, puede constituirse en pauta única y suficiente.

Con todo, el control de calidad de estos recursos informativos es tarea imprescindible para una sociedad en la que el acceso a información es el paso previo y necesario para la producción de ciencia y tecnología. Y de la calidad de las fuentes que la proporcionan dependerá, de modo directo,

no sólo el desarrollo de la comunidad científica, sino también el de la sociedad en todo su conjunto.

BIBLIOGRAFÍA

- ALTUNA, B. ACEBES, L. MORENO, A. DOMÍNGUEZ, P. Estudio de tipología de errores en bases de datos bibliográficas nacionales para el proyecto DELICAT/KSYSERROR, en *V Jornadas Españolas de Documentación Automatizada, Cáceres, 17, 18 y 19 de octubre de 1996*. Cáceres: Universidad de Extremadura, 1996, vol. 2, p. 835-838.
- AMSTRONG, C. J.; MEDAWAR, K. *Investigation into the quality of databases in general use in the U.K. British Library Project*, october 1994-april 1995.
- BROCKMAN, J. Just another management fad? The implications of TQM for library and information services. *ASLIB Proceedings*, 1992, vol. 44, n 7/8, p. 283-288.
- CHU, C. M.; ALJIFERUQUE, I. Quality of indexing library and information science databases. *Online Review*, 1989, vol. 13, n 1, p. 11-35.
- DUARTE, B. Indicadores como instrumentos de evaluación de los servicios bibliotecarios: aspectos metodológicos. *Boletín de la ANABAD*, 1995, v. XLV, n 1, p. 95-106.
- DUFLOS, A. *La qualité en marche dans les services d'information*. Paris: ADBS, 1995.
- La qualité: enquête chez les producteurs de services électroniques d'information. *3^e Forum La qualité des services d'information documentaire*. Paris, diciembre 1994, n 25.
- GARCÍA-MORALES HUIDOBRO, E. Aportaciones de la gestión de calidad a bibliotecas y servicios de documentación. *Revista Española de Documentación Científica*, 1995, vol. 18, n^o 1, p. 1-
- GUILBERT, G. Qualité des données et des services associés. *Journée d'étude ADBS-GFII. Information sur les entreprises: comment choisir?* Document 8, Paris, 1995.
- HARRY, V.; OPPENHEIM, Ch. Evaluations of electronic databases. Part I: Criteria for testing CD-ROM products. *Online & CD-ROM Review*, 1993, vol. 17, n 4, p. 211-222.
- HARRY, V.; OPPENHEIM, Ch. Evaluations of electronic databases. Part II: Testing CD-ROM products. *Online & CD-ROM Review*, 1993, vol. 17, n 6, p. 339-347.
- HERRERO SOLANA, V. La calidad total en bases de datos españolas: estudio de la tasa de error en las bases del CSIC. *Revista Española de Documentación Científica*, 1997, vol. 20, n 4, p. 409-416.
- INFORMATION Market Observatory (IMO). *The quality of electronic information products and services*. Luxembourg, September 1995. IMO Working Paper 95/4, p. 1-16.
- JACSÓ, P. Searching for skeletons in the database cupboard. Part I: errors of omission. *Database*, 1993, vol. 16, n 1, p. 67-76.
- Searching for skeletons in the database cupboard. Part II: errors of commission. *Database*, 1993, vol. 16, n 2, p. 77-82.
- KHUN, P., et al. Criteria of quality assesment for scientific databases. *Journal of Chemical Information and Computer Sciences*, 1994, vol. 34, n 3, p. 517-519.
- KUME, H. *Herramientas estadísticas básicas para el mejoramiento de la calidad*. Madrid: Paramon, 1996.
- LESTER, D. E. *The impact of quality mngement on the information sector: a study of case histories*. EUSIDIC, 1994.

- MEDAWARD, K. Database quality: a literature review of the past and a plan for the future. *Program*, 1995, vol. 29, n 3, p. 23-29.
- MOSCOSO, P. Pautas para evaluar bases de datos en CD-ROM. *Revista General de Información y Documentación*, 1997, vol. 7, n 1, p. 187-204.
- ORTEGA DE LORENZO-CÁCERES; P.; BONAL ZAZO, J. L. Indicadores para el control de calidad de bases de datos bibliográficas españolas. En: *V Jornadas Españolas de Documentación Automatizada, Cáceres 17-19 de octubre de 1996*. Cáceres: Universidad de Extremadura: ABADMEX, 1996, vol. 1; p. 503-512.
- O'NEILL, E. T.; ROGERS, S.; OSKIN, W. M. Characteristics of duplicate records in OCLCs Online Union Catalogue. *Library Resources and Technical Services*, 1993, vol. 37, n 1, p. 60.
- O'NEILL, E.T.; VIZINE-GOETZ, D. Quality control in online databases. *Annual Review of Information Science and Technology*, 1988, vol. 23, p. 125-156.
- O'NEILL, E.T.; VIZINE-GOETZ, D. The impact of spelling errors on databases and indexes. *Proceedings of the National Online Meeting*, New York, 1987, May 9-11, p. 313-320.
- PACIOS LOZANO, A. R. Sistemas de calidad total en bibliotecas y unidades de información. *Boletín ANABAD*, 1995, n 2, p. 107-116, p. 109.
- PALMA VILLALIN, M. V. Técnicas y métodos para mejorar la calidad de la indización y su recuperación en bases de datos documentales de ciencias sociales y humanidades. En: *5ª Jornades Catalanes de Documentació: biblioteques, centres de documentació i serveis de informació, 25-27 d'octubre de 1995, Palau de Congressos de Barcelona*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya; Societat Catalana de Documentació i Informació, 1995, p.
- SCHWARTZ, C. Evaluating CD-ROM products: Yet another list. *CD-ROM Professional*, 1993, vol. 4, n 1, p. 87-105.
- SPINAK, E. Errores ortográficos en el ingreso en bases de datos. *Revista Española de Documentación Científica*, 1995, vol. 18, n 3, p. 307-319.
- TALADRIZ MAS, M. Control de calidad y costes de los servicios de información. *Revista Española de Documentación Científica*, 1994, vol. 17, n 4, p. 450-459.
- TONEY, V. Cleanup and the duplication of an International Bibliographic Database. *Information Technology and Libraries*, 1992, vol. 11, p. 21-23.
- WHITE, H. D.; GRIFITH, B. C. Quality of indexing in online databases. *Information Processing and Management*, 1987, vol. 23, p. 211-214.
- WULFF BARREIRO, E. Calidad de la indización en bases de datos de tamaño cualesquiera. *Boletín de la Asociación Andaluza de Bibliotecarios*, 1993, vol. 9, n 31, p. 41-45.