

Métodos de evaluación de calidad de catálogos automatizados: análisis comparativo

M.^a del Pilar Ortego de Lorenzo-Cáceres

C.e. portego@alcazaba.unex.es

José Luis Bonal Zazo

C.e. jlbonal@alcazaba.unex.es

Universidad de Extremadura. Facultad de Biblioteconomía y Documentación

Resumen: Se describen las características de los principales métodos de evaluación de calidad de catálogos automatizados, con los siguientes objetivos: presentar, de forma sintética, los métodos de evaluación de calidad más comunes, proponer unas pautas de aplicación y uso de los distintos sistemas de evaluación y analizar las diferencias existentes entre los métodos presentados, con el fin de establecer cuál es el sistema de medida más adecuado para cada tipo de análisis. Los métodos de evaluación analizados son: 1) técnicas de muestreo; 2) tests de suciedad; 3) análisis transaccionales; 4) técnicas de simulación y experimentación; y 5) análisis de palabras de baja frecuencia.

1. Introducción

La aparición, en los últimos años, de numerosos métodos de evaluación de calidad de bases de datos bibliográficas, aunque ha contribuido a mejorar los sistemas de análisis, ha acarreado, al mismo tiempo, una complicación notable de los procedimientos de trabajo: mientras que algunos métodos de evaluación sólo son apropiados para determinados tipos de bases de datos, otros son útiles para el análisis de aspectos concretos o partes delimitadas de una base de datos, de un registro o, incluso, de un campo; y otros, en cambio, permiten un análisis general de todos los aspectos, todos los problemas y todos los elementos.

Con el objetivo de contribuir a delimitar el uso de algunos de estos métodos se han examinado, de forma sintética, las características y utilidad de cinco métodos de análisis de bases de datos: las técnicas de muestreo, los tests de suciedad, los análisis transaccionales, las técnicas de experimentación, y el análisis de palabras de baja frecuencia. De cada uno de estos métodos se han analizado los siguientes aspectos:

1. Tipo de base de datos a la que se pueden aplicar los métodos citados (de carácter bibliográfico, o de carácter documental).
2. Grado de análisis. Diferenciando si el sistema de evaluación permite un análisis global de la base de datos, o, por el contrario, simplemente permite un análisis de los registros o de parte de un registro.

3. Utilidad para medir los errores de consistencia (errores de interpretación y aplicación de las normas).
4. Utilidad para medir los errores de exactitud o de precisión (errores en los caracteres).
5. Grado de calidad y representatividad global de cada sistema de evaluación.

Tabla comparativa de análisis de métodos de evaluación

	Sistema utilizado				
	1	2	3	4	5
<i>Tipo de Base de datos</i>					
Bibliográfica	•	•	•	•	•
Documental	•	•	•	•	•
<i>Grado de análisis</i>					
Base de datos			•	•	
Registro	•	•	•	•	•
<i>Errores de consistencia</i>					
Errores de interpretación normativa	•		•	•	
Omisión de datos mínimos obligatorios	•		•	•	
Errores de términos de indización	•		•	•	
Errores en formatos electrónicos	•		•	•	
<i>Errores de exactitud o precisión</i>					
Errores gramaticales					
Ortográficos	•	•	•	•	•
Sintácticos	•		•	•	
Semánticos	•		•	•	
Errores tipográficos o mecanográficos					
Permutación	•	•	•	•	•
Omisión	•	•	•	•	•
Sustitución	•	•	•	•	•
Repetición de letras	•	•	•	•	•
Inserción de letras	•	•	•	•	•
Inserción de espacios	•	•	•	•	•
Supresión de espacios	•	•	•	•	•

Muestreo (1); tests de suciedad (2); análisis transaccional (3); técnicas de experimentación (4); análisis de palabras de baja frecuencia (5).

2. Las técnicas de muestreo

La utilización del muestreo estadístico para el análisis de la calidad de las bases de datos es uno de los métodos de evaluación usados con más frecuencia, tanto de forma independiente como combinada con cualquiera de los otros métodos ex-

puestos¹. La principal ventaja de este procedimiento es que se basa en técnicas aplicadas comúnmente por otras disciplinas y, por tanto, con un alto grado de experimentación y uso, de forma que su aplicación está ya comprobada.

Básicamente, el muestreo estadístico para la selección de un número representativo de registros bibliográficos se desarrolla en tres etapas: 1) la determinación de la muestra objeto de análisis; 2) la recogida de datos y 3) el análisis de datos.

Por lo que respecta a la determinación de la muestra, hay que tener en cuenta que dado que, comúnmente, la población objeto de selección en los catálogos es una población de carácter infinito (superior a las 10.000 unidades), es posible aplicar medidas de cálculo del tamaño muestral generales, usadas para estimar proporciones representativas para poblaciones infinitas. En cualquier caso, en varios estudios realizados basados en análisis de muestras, el tamaño muestral varía notablemente, oscilando entre muestras de menos de 200 registros y muestras de más de 2000. Una vez decidida la cantidad de registros a seleccionar es necesario establecer el sistema de selección. El procedimiento de selección más objetivo es la extracción de registros al azar (aplicando cualquier sistema de progresión numérica, a través de las iniciales del contenido de los campos, o por cualquier otro procedimiento aleatorio); sin embargo, dependiendo de las características del estudio, es posible utilizar determinados criterios de selección que contribuyan a la creación de muestras representativas, tales como:

- la lengua de los documentos;
- el tipo de documento;
- la originalidad de los registros;
- la profundidad en el nivel de descripción bibliográfica o documental;
- la fecha de los documentos o de creación de los registros;
- la materia de los documentos;
- la edición de las obras;
- las reglas o normas empleadas para la descripción

La segunda etapa del análisis muestral se desarrolla en dos fases: una fase de sistematización, en la que es preciso establecer un procedimiento para facilitar el análisis de los registros seleccionados en la muestra; y una segunda fase, de codificación de errores, con la finalidad de permitir identificar con precisión el tipo de error, así como indicar con exactitud su localización. En la mayoría de las ocasiones, no obstante, no basta con recoger los registros, sino que es necesario, además, recoger información bibliográfica procedente de otras bases de datos, con la finalidad de contrastar la información objeto de análisis.

En su conjunto, las técnicas de muestreo aplicadas al estudio de las bases de datos bibliográficas presentan importantes ventajas. Es posible su aplicación a cualquier tipo de base de datos, para el análisis de cualquier elemento de los registros, y

¹ Realmente, cualquiera de los métodos expuestos utiliza técnicas estadísticas, más o menos complejas, para el análisis de los resultados obtenidos.

para el estudio, además, de cualquier tipo de error (de exactitud o de consistencia), todo ello, con un porcentaje elevado de fiabilidad. El único problema destacable es su limitación para la evaluación general de la base de datos.

3. Tests de suciedad

Entre los distintos tipos de tests existentes para la evaluación de la calidad de bases de datos (de consistencia, de duplicación, de sintaxis...) destacan los de suciedad. Se trata de tests basados en la búsqueda, en una base de datos, de un número de términos representativos escritos deliberadamente de forma errónea, con el fin de analizar su frecuencia de aparición y, sobre todo, su relación proporcional con el número de palabras escritas correctamente. Se trata de un método de evaluación relativamente reciente, ya que los primeros trabajos que aplican tests de suciedad con fines de análisis de calidad, datan de principios de los años noventa².

Aunque existen diferentes tipos de tests de suciedad, todos ellos se basan en el modelo expuesto, con pequeñas variantes orientadas a aumentar la representación de los resultados obtenidos y a precisar los métodos de análisis y exposición³. En cualquiera de los tipos de tests de suciedad, no obstante, la metodología de aplicación es similar, desarrollándose en cuatro operaciones:

1. selección de la serie de términos erróneos. El número de términos usados por la mayoría de los modelos oscila en torno a 10, no obstante, algunos modelos truncan los términos, o buscan todas las variantes de error del término erróneo, con el fin de obtener resultados más fidedignos, en cualquier caso, debe tratarse de palabras significativas y con alto riesgo de error.
2. Búsqueda en la base de datos del número de términos erróneos seleccionados y de sus variantes, y comprobación, en su contexto, de que efectivamente se trata de un término erróneo.
3. Reducción de los datos totales a los porcentajes adecuados, con el fin de facilitar el análisis comparativo con otras bases de datos.
4. Análisis de resultados (relación entre palabras mal escritas y palabras bien escritas, elementos de la descripción donde son más frecuentes los errores, análisis comparativos, etc.).

En general, los tests de suciedad son apropiados para el análisis de cualquier tipo de base de datos, aunque sólo a nivel de registro y únicamente, debido a su

² DWYER, Jim. «The cataloger's "invisible college" at work: the case of the dirty database test». *Cataloging & Classification Quarterly*. 1991, vol. 14, n.º 1, p. 75-82; CAHN, Pamela. «Testing database quality». *Database*. 1994.; NICHOLS RANDAL, Barbara. «Spelling errors in the database shadow or substance». *Library resources & technical services*. 1999, vol. 43, n.º 3, p. 161-169.

³ Un análisis de éstos puede encontrarse en nuestro trabajo: ORTEGO DE LORENZO-CÁCERES, M.ª del Pilar; BONAL ZAZO, José Luis. «Evaluación de calidad de catálogos mediante el uso de tests de suciedad». En *Jornadas Bibliotecarias de Andalucía* (11, 2000, Sevilla). *Profesionales para una sociedad bien informada*. Málaga: Asociación Andaluza de Bibliotecarios, 2000, p. 229-236.

carácter, son válidos para medir errores gramaticales de carácter ortográfico y errores mecanográficos de cualquier tipo, pero no errores de consistencia. Su nivel de fiabilidad, en cambio, es reducido, debido a que, en definitiva, se basan en métodos de probabilidad aleatorios, con un número de palabras muy pequeño, con lo cual cualquier factor no controlado puede modificar notablemente los resultados, falseando el grado de calidad.

4. El análisis transaccional

El análisis transaccional es un procedimiento de investigación basado en el análisis de las actividades desarrolladas por una persona al utilizar un ordenador, que se encuentran grabadas en la memoria de éste. La técnica del análisis transaccional consiste en analizar pormenorizadamente todas las operaciones o «transacciones» que han realizado las personas que han manipulado un sistema informático. En el ámbito de la Biblioteconomía y la Documentación esta técnica de investigación y análisis se ha orientado fundamentalmente al estudio del comportamiento de los usuarios ante el ordenador, con la finalidad de mejorar los sistemas de búsqueda y recuperación de la información (a partir del examen de cuestiones tales como el número y tipo de búsquedas realizadas, el sistema de recuperación usado, los errores cometidos en la búsqueda, el sistema de corrección de éstos, el uso de los mecanismos de ayuda, etc.⁴).

La utilización del análisis transaccional para la evaluación de la calidad de las bases de datos ha sido limitada, debido, en buena parte, a las dificultades existentes para la recogida de la información, dado que el análisis se basa en el estudio de las veces que los registros de una base de datos son editados por los analistas para hacer algún tipo de corrección en su contenido, de manera que el estudio de las correcciones realizadas permita determinar qué tipo de errores son los más habituales⁵. No obstante, no es un método de análisis puro, debido a que su aplicación debe ser realizada con técnicas de muestreo (selección de las transacciones realizadas durante un periodo de tiempo establecido, o realizadas por un grupo determinado de analistas, etc.).

En general, se trata de un método de análisis adecuado, sobre todo, para la evaluación de catálogos colectivos o bases de datos bibliográficas alimentadas por un número elevado de catalogadores. A diferencia de otros sistemas de evaluación, presenta una ventaja notable: permite no sólo evaluar todos los elementos de cada registro, sino también la base de datos completa, debido a que, frecuentemente, las

⁴ Un buen ejemplo del uso del análisis transaccional para la evaluación del comportamiento de los usuarios puede encontrarse en MILLSAP, Larry; FERL, Terry Ellen. «Search patterns of remote users: an analysis of OPAC transaction logs». *Information technology and libraries*. 1993, sept., vol. 12, n.º 3, p. 321-343.

⁵ CHAPMAN, Ann. «Up to standard? A study of the quality of records in a shared cataloguing database». *Journal of librarianship and information science*. 1994, diciembre, vol. 26, n.º 4, p. 201-210.

transacciones implican la edición y eliminación de registros duplicados, el establecimiento de relaciones entre registros y la revisión de otros aspectos comunes al conjunto de la base de datos. Igualmente, aunque se trata de un método que permite evaluar los errores de consistencia y los de exactitud, es, sobre todo, apropiado para el análisis de los primeros, ya que, comúnmente, la edición de registros se realizará para solucionar problemas relacionados con documentos que presentan dificultades de interpretación o de utilización de las reglas de catalogación, listas de indización o tablas de clasificación. Aunque es un sistema de evaluación poco objetivo (la edición de registros es realizada por los analistas, en función de su criterio particular), la representatividad del tipo de problemas que pueden ser encontrados en el análisis documental y que, en consecuencia, pueden encontrarse en la base de datos, es alta.

5. Técnicas de experimentación

A diferencia del resto de los métodos de evaluación expuestos, en los cuales el análisis se efectúa sobre datos reales, la experimentación es una técnica basada en datos ficticios, en la cual se controlan con precisión todas las condiciones que pueden influir en el fenómeno estudiado, y se comprueba y contrasta éste bajo todas las condiciones controladas. Es decir, se crea un modelo de análisis en el que se tienen en cuenta todos los factores que pueden interferir en los resultados (por ejemplo, se crea una base de datos en la que se controlan todos los factores que pueden afectar a la recuperación de los documentos).

Las fases de una experimentación en Biblioteconomía variarán, dependiendo de las características de la experimentación, pero debe contar, al menos, con las siguientes⁶:

1. formulación de la hipótesis de partida.
2. definición de las variables.
3. control de los factores que pueden influir en los resultados de la experimentación.
4. diseño de la experimentación, estableciendo las pruebas necesarias para confirmar o rechazar la hipótesis formulada.
5. recogida de datos y análisis y expresión de los mismos.

En el campo de la Biblioteconomía, la experimentación se usa con varios fines: probar nuevas herramientas y nuevas técnicas, analizar el uso de las colecciones, analizar las causas de determinados problemas, etc. Particularmente, por lo que respecta a su aplicación al análisis de bases de datos bibliográficas, su utilización presenta dos características notables. En primer lugar, su aplicación es más preventiva que correctiva: los estudios realizados que han aplicado técnicas de

⁶ GARCÍA ROLDAN, José Luis. *Cómo elaborar un proyecto de investigación*. Alicante: Universidad, 1995.

análisis experimental lo han hecho con la finalidad de prevenir posibles problemas, no para detectar los existentes en bases de datos reales, aunque a partir del primer tipo de análisis se pueda llegar al segundo. Por otra parte, la mayoría de éstos son estudios orientados al análisis de las bases de datos desde el punto de vista de la recuperación (buena parte de los estudios realizados son pruebas hechas con usuarios) y no desde el punto de vista del almacenamiento de la información (perspectiva en la que se encuentra la evaluación de bases de datos)⁷. En cualquier caso, las técnicas aplicables para la «evaluación preventiva» de una base de datos, estarán basadas en estudios de simulación de situaciones reales, o de experimentación controlada.

De forma global, las técnicas de experimentación, permitirán el análisis de cualquier tipo de bases de datos, dependiendo del diseño de la experimentación. El diseño de la experimentación influirá también en los elementos que pueden ser objeto de análisis: aunque potencialmente, un estudio experimental puede permitir el análisis tanto de registros, de forma independiente como en conjunto. Del mismo modo, un estudio experimental puede permitir medir tanto los errores de consistencia como los de precisión, aunque siempre será con carácter preventivo, desde el punto de vista experimental, no real. Pese a esta circunstancia, se trata de un método con un porcentaje de fiabilidad extraordinariamente elevado, ya que, generalmente, ofrece resultados válidos sobre el total de los elementos analizados, con todas las variables completamente controladas.

6. Análisis de palabras de baja frecuencia⁸

El análisis de palabras de baja frecuencia es un procedimiento de evaluación basado en la idea de que el número de palabras erróneas que se encuentran en una base de datos es reducido (diversos estudios que han utilizado otros métodos han demostrado que el porcentaje de términos erróneos es, frecuentemente, inferior al 1%); por esta razón, es previsible que la mayor parte de las palabras erróneas que se encuentren en una base de datos puedan ser localizadas generando un índice de

⁷ Una revisión bibliográfica de los estudios basados en pruebas experimentales puede encontrarse en LARGE, Andrew; BEHESHTI, Jamshid. «OPAC's: a research review». *Library and information Science Research*. 1997, vol. 19, n.º 2, p. 116-118. Las técnicas presentadas en este trabajo varían desde la creación de bases de datos experimentales para realizar estudios de simulación incluyendo puntos de acceso suplementarios con el fin de mejorar el acceso, hasta modelos matemáticos de simulación de sistemas de recuperación de información, pasando por pruebas de búsquedas con diferentes tipos de usuarios.

Es interesante, asimismo, el modelo experimental propuesto por Lisa Romero en 1992, en el cual la autora seleccionó a 20 catalogadores, con el fin de que catalogaran las mismas obras (10 en total), a fin de determinar cuáles eran las áreas y los elementos que, potencialmente, podrían ocasionar más problemas de almacenamiento. Cfr. ROMERO, Lisa. «Original cataloging errors: a comparison on errors found in entry-level cataloging with errors found in OCLC and RLIN». *Technical services quarterly*. 1994, vol. 12, n.º 2, p. 13-27.

⁸ Véase SPINAK, Ernesto. «Errores ortográficos en el ingreso en bases de datos». *Revista española de documentación científica*. 1995, vol. 18, n.º 3, p. 312-313.

las palabras de la misma, ordenado por frecuencias de aparición, de tal forma que en la parte del índice donde se localicen las palabras con menor frecuencia se encontrarán la mayor parte de los errores, mientras que, por el contrario, a medida que aumenta la frecuencia de aparición de las palabras disminuye el riesgo de error⁹.

Se trata de un método usado no sólo para la evaluación de la calidad de bases de datos, sino también para la corrección, al menos ortográfica de una buena parte de los términos de la misma mediante su aplicación por procedimientos automatizados¹⁰.

No obstante, pese a sus ventajas en un contexto automatizado, presenta numerosas limitaciones. En primer lugar, su grado de análisis es reducido, debido a que no permite analizar la calidad del conjunto de la base de datos, sino sólo de determinados elementos de cada registro fuera de su situación original, de tal forma que su análisis es más complejo. En segundo lugar, se trata de un sistema que no permite analizar los errores de consistencia, ni los errores de precisión gramaticales de carácter semántico o sintáctico; sí resulta, en cambio, un método apropiado para el análisis de errores ortográficos, así como de errores mecanográficos de todos los tipos posibles (permutación, omisión, sustitución, repetición de letras, inserción de letras, inserción de espacios en blanco o supresión de espacios). De este modo, este sistema resulta apropiado, exclusivamente, para el análisis de los errores ortográficos y mecanográficos, proporcionando, sólo en éstos, un elevado grado de fiabilidad y de representación.

7. Conclusión

Aunque de todos los métodos apuntados, es la evaluación mediante el uso de técnicas de muestreo el procedimiento más utilizado y más apropiado para hacer una evolución integral y sistemática de las bases de dato, todos los métodos expuestos, son, por su facilidad de aplicación útiles para la evaluación de aspectos concretos. Así, técnicas tales como los tests de suciedad o el análisis de palabras de baja frecuencia resultan de gran utilidad para el estudio de los errores de precisión; mientras que otros métodos, tales como el análisis transaccional son apropiados para el examen de los errores de consistencia.

⁹ *Ibidem*

¹⁰ NEILL, Edward T.; VIZINE-GOETZ, Diane. «Quality control in online databases». *Annual Review of Information Science and Technology*. 1988, vol. 23, p. 125-156.