

Indicadores para evaluar el vocabulario y la estructura sistemática de un tesoro

Ana M. Martínez^{1,2}, Cristina A. Ristuccia², Edgardo A. Stubbs², Julia C. Valdez²,
Viviana L. Gamba², Paola V. Mendes², Carolina Unzurrunzaga², María L. Caminotti²

¹Instituto de Investigaciones en Humanidades y Ciencias Sociales (UNLP-CONICET). ²Departamento de Bibliotecología, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. E-mail: ammarti@speedy.com.ar

Resumen: Se proponen siete indicadores para evaluar la calidad del vocabulario y la estructura sistemática del tesoro, considerando que las especificaciones incluidas en las normas para su diseño y desarrollo, así como los principios de la teoría de la clasificación son los requisitos de calidad que deben alcanzar las características inherentes de los tesoros. Las características que no cumplen con dichos requisitos se consideran no conformidades. Los indicadores fueron probados en tres tesoros en español (DeCS, Spines y Tesoro de la Unesco). Los resultados fueron: a) Términos preferentes que representan más de un concepto (ideal 0): DeCS 10,8%, Spines 7,3%, Unesco 10,9%. b) Media de palabras por término preferente (ideal ≤ 3): DeCS $2,3 \pm 1,6$, Spines $1,9 \pm 0,9$, Unesco $2,0 \pm 1,1$. c) Términos preferentes con formas gramaticales no aceptadas (ideal 0): DeCS 19,8%, Spines 10,0%, Unesco 15,8%. d) Términos preferentes con formas de singular y plural no aceptadas (ideal 0): DeCS 4,7%, Spines 2,2%, Unesco 40,5%. e) Términos preferentes que carecen de relación jerárquica (ideal 0): DeCS 0, Spines 1,4% y Unesco 1,1%. f) Términos preferentes con dos o más términos genéricos (ideal 0): DeCS 53,8%, Spines 14,9%, Unesco 0. g) Términos preferentes con un solo término específico (ideal 0): DeCS 13,5%, Spines 3,5%, Unesco 10,3%. Se concluye que estos indicadores cumplen con los criterios de un buen indicador y pueden ser útiles para la evaluación de la calidad del tesoro durante el proceso de diseño y desarrollo.

Introducción

El diseño y desarrollo de tesoros ha estado normalizado por más de 35 años. Sin embargo, la experiencia demuestra que en muchos casos los tesoros no cumplen estrictamente con las normas, por lo que es importante evaluar la calidad de los mismos.

La evaluación de los tesauros ha sido tratada por diversos autores, durante largo tiempo, con distintas técnicas. Gil Leiva (2008) sintetiza estos aportes en dos tipos de evaluación:

- Evaluación intrínseca: su objeto de análisis es el tesoro propiamente dicho. Esta evaluación puede ser, a su vez, de dos tipos: a) cualitativa (Lancaster, 2002) la denomina superficial), que evalúa si son adecuados aspectos tales como la introducción, el uso de calificadores y notas de alcance, la reciprocidad de las relaciones, las formas gramaticales de los términos preferentes, etc. y b) cuantitativa (Lancaster, 2002) la denomina estadística), en la que se pueden aplicar indicadores.

- Evaluación extrínseca: su objeto de análisis es el comportamiento del tesoro en la indización y la recuperación de información.

Desde otro punto de vista, se puede afirmar que hay tres etapas en las que puede ser evaluado un tesoro: la etapa de diseño y desarrollo, la etapa de indización y la etapa de recuperación.

Para la etapa de indización Aitchison et al. (2000) y Lancaster (2002) coinciden en recomendar un período de prueba durante el cual el tesoro es utilizado por distintos indizadores, en distintas unidades de información, con el fin de obtener la retroalimentación necesaria para realizar ajustes y correcciones en el tesoro. Por su parte, Owens y Cochrane (2004) han propuesto un algoritmo para evaluar el tesoro tanto en la indización como en la recuperación.

Para la recuperación las mediciones más conocidas son las tasas de precisión y de acierto propuestas por Claverdon y Mills (1985). En el resultado de estas mediciones intervienen distintas variables como la exhaustividad y la especificidad de la indización, la interfaz usuario-sistema, la experiencia de búsqueda de los usuarios finales, etc., pero también se han utilizado para evaluar específicamente la recuperación por términos preferentes, ya sea de un tesoro o de otros vocabularios controlados. La tasa de precisión ha sido incorporada como indicador en la norma IRAM 11620:2007 (Instituto Argentino de Normalización, 2007), mientras que la tasa de acierto, hoy en día, puede tener una importancia especial para evaluar la búsqueda expandida (entendida como la búsqueda simultánea en varias bases de datos y la Web con los mismos términos), aunque otros

autores han propuesto formas alternativas de evaluar este tipo de búsqueda (International Organization for Standardization, 2000).

La evaluación del tesoro en las etapas de indización y recuperación coincide con la evaluación extrínseca de Gil Leiva (2008) y escapan al propósito del presente trabajo, que se centra en la evaluación de distintos aspectos del tesoro, mediante el uso de indicadores (evaluación intrínseca estadística), durante la etapa de diseño y desarrollo, es decir antes de que el tesoro sea utilizado por indizadores y usuarios finales.

Por otro lado, a norma ISO 9000:2000 define *calidad* como “el grado en el que un conjunto de características inherentes cumple con los requisitos”. A su vez, una *característica inherente* es un “rasgo diferenciador” que “existe en algo [producto, proceso, sistema], especialmente como una característica permanente”. Por último, *requisito* se define como “la necesidad o expectativa establecida, generalmente implícita u obligatoria”. De modo que si las características inherentes cumplen con los requisitos, la calidad es alta, de lo contrario la calidad es baja. Una característica inherente que no cumple con los requisitos correspondientes es una no conformidad. Para medir las no conformidades de un tesoro es conveniente contar con indicadores (International Organization for Standardization, 2000). Según la norma IRAM 11620:2007, un indicador es "una expresión (que puede ser numérica, simbólica o verbal) utilizada para caracterizar actividades (sucesos, objetos, personas) en términos cuantitativos y cualitativos, con el propósito de evaluar el valor de las actividades caracterizadas y el método asociado". Los indicadores son así herramientas que proveen un valor de referencia a partir del cual se puede establecer una comparación (Instituto Argentino de Normalización, 2007). Orr (1973) estableció los criterios que definen un buen indicador, que luego fueron incorporadas en la ya mencionada norma IRAM 11620:2007. Éstos son los siguientes:

- Contenido informativo: el indicador tiene que ser informativo como herramienta para medir una actividad e identificar tanto logros como problemas y deficiencias, con el objetivo de poder actuar para remediarlos.
- Confiabilidad: el indicador debe producir el mismo resultado cuando se utiliza repetidamente bajo las mismas condiciones o circunstancias.
- Validez: el indicador debe medir lo que se quiere medir.

- Adecuación: el indicador debe ser acorde con el objetivo para el que se ha formulado. Las unidades, escalas y operaciones necesarias para implementar el proceso de medición deben ser apropiadas.
- Aplicabilidad: el indicador debe ser práctico, es decir debe utilizar datos que resulten accesibles, con un esfuerzo razonable en términos de tiempo y costos operativos.
- Comparabilidad: el indicador debe permitir que la calidad de dos o más actividades pueda ser comparada. En el caso del presente trabajo, debe permitir la comparación de la calidad de dos o más tesauros.

La experiencia más conocida para la evaluación de tesauros usando indicadores es la realizada por el Bureau Marcel Van Dijk (BMD) de Bruselas, que en 1976 llevó a cabo un estudio sobre 10 tesauros monolingües y 10 multilingües (Commission des Communautés Européennes, 1976). El BMD propuso varios indicadores que, al igual que el presente trabajo, se orientaban a la etapa de diseño y desarrollo, ya que incluso recomendaban su inclusión en la norma internacional. Otros autores han aplicado o recomendado estos indicadores (Alvaro Bermejo et al., 1989; Gil Leiva, 2008; Gil Urdiciain, 1998), aunque Lancaster (2002) llegó a la conclusión de que si bien algunos eran ingeniosos, otros resultaban triviales e incluso arbitrarios. Por otro lado, hay que tener en cuenta que el BMD basó su evaluación en la primera edición de la norma ISO 2788:1974, que fue actualizada en 1986 y se encuentra actualmente en revisión (International Organization for Standardization, 1986).

En consecuencia, el propósito del presente trabajo es proponer indicadores que permitan medir las no conformidades en las formas gramaticales y la estructura sistemática de los tesauros, respecto tanto a las normas para su diseño y desarrollo, como a los principios de la teoría de la clasificación.

Materiales y métodos

En primer lugar, aquí se considera que los requisitos son las especificaciones establecidas por las normas para el diseño y desarrollo de tesauros, así como los principios de la teoría de la clasificación (para algunos aspectos de la estructura sistemática), mientras que las características inherentes son los rasgos que presentan los tesauros evaluados. Las normas tomadas en cuenta para identificar los requisitos son las siguientes:

- ISO 2788:1986 para tesauros monolingües. Esta norma está vigente, pero en revisión, ya que una nueva norma ISO 25964 se encuentra en preparación y reemplazará a la anterior (International Organization for Standardization, 1986).
- ANSI/NISO Z39.19:2005. Es la cuarta edición de la norma estadounidense que ha sido modificada para incluir otros sistemas de organización del conocimiento. Aquí se toman en cuenta solamente las especificaciones pertinentes a los tesauros (National Information Standards Organization, 2005).
- BS 8723-2:2005. Es la tercera edición de la norma británica y al igual que la estadounidense incluye otros sistemas de organización de conocimiento. El tesoro está tratado en la Parte 2, que ha sido tomada en cuenta para este trabajo (BSI Group, 2005).

Se proponen los siguientes indicadores,

a) Porcentaje de términos preferentes que designan más de un concepto. Se define como $A/B*100$, donde A es el número de términos preferentes que designan más de un concepto y B es el número total de términos preferentes.

b) Media de palabra por término preferente. Se define como A/B , donde A es el número de palabras de los términos preferentes y B es el número total de términos preferentes. Se expresa como $media \pm DE$ ($media \pm$ un desvío estándar). El valor de referencia es $2,0 \pm 1,0$, es decir un rango promedio entre 1 y 3 palabras por término preferente.

c) Porcentaje de términos preferentes con formas gramaticales no aceptadas. Se define como $A/B*100$, donde A es el número de términos preferentes que presentan formas gramaticales distintas a las especificadas por las normas y B es el número total de términos preferentes. El valor de referencia debería ser 0, pero como las normas admiten excepciones, se debe asumir que mientras menor sea este porcentaje, mayor será la calidad de los términos preferentes.

d) 4 Porcentaje de términos preferentes con formas de singular y plural no aceptadas. Se define como $A/B*100$, donde A es el número de términos preferentes que presentan formas de singular o plural distintas a las especificadas por las normas y B es el número total de términos preferentes. El valor de referencia es 0, ya que en este caso no hay excepciones.

e) Porcentaje de términos preferentes que carecen de relación jerárquica (términos huérfanos). Se define como $A/B*100$, donde A es el número de términos preferentes que carecen de relación jerárquica y B es el número total de términos preferentes. El valor de referencia es 0.

f) Porcentaje de términos preferentes con dos o más términos genéricos (TG). Se define como $A/B*100$, donde A es la suma de términos preferentes que poseen dos o más TG y B es el número total de términos preferentes. El valor ideal debería ser 0, pero como las normas admiten excepciones, se debe asumir que mientras menor sea este porcentaje, mayor será la calidad de las relaciones jerárquicas en relación a los términos genéricos.

g) Porcentaje de términos preferentes con un solo término específico (TE). Se define como $A/B*100$, donde A es la suma de términos preferentes que poseen un solo TE y B es el número total de términos preferentes. Aunque las normas no lo mencionan, este indicador responde al Canon de Diferenciación postulado por Ranganathan, según el cual cuando se aplica un principio de división a un concepto deben resultar, al menos, dos conceptos más específicos (Mills, 2004; Ranganathan, 1967; Spiteri, 1998; Vickery, 1960). El valor de referencia es 0.

Para probar estos indicadores, se llevó a cabo una prueba de aplicación en una muestra de términos preferentes tomados de tres tesauros en español:

- Descriptores en ciencias de la salud (DeCS): especializado en ciencias de la salud. Es una adaptación del MeSH, desarrollado por BIREME-Centro Latinoamericano y del Caribe en Ciencias de la Salud (BIREME, 2008).
- Tesoro Spines (Spines): es un tesoro especializado en ciencia y tecnología, desarrollado por la Unesco (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 1988).
- Tesoro de la Unesco (Unesco): es también un tesoro desarrollado por la Unesco sobre educación, ciencia y cultura (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2007).

La prueba de aplicación se realizó sobre una muestra al azar de cada uno de los tesauros mencionados (nivel de confianza 95%, intervalo de confianza 5), resultando 379 términos preferentes para DeCS, 370 para Spines y 368 para Unesco. No se incluyeron nombres propios en la muestra.

Resultados y discusión

Los resultados de la prueba realizada en los tres tesauros se muestra en la Tabla 1.

El primer indicador (a) revela que cerca del 10% de los términos preferentes representan más de un concepto, cuando el valor de referencia es 0. Esta característica es propia de otros vocabularios, como las listas de epígrafes, pero no del tesauro.

Tabla 1. Indicadores aplicados a una muestra de tres tesauros

Indicadores	Valor de referencia	DeCs	Spines	Unesco
Términos preferentes que representan más de un concepto	0	10,8%	7,3%	10,9%
Media de palabras por término preferente	≤ 3	2,3 \pm 1,6	1,9 \pm 0,9	2,0 \pm 1,1
Términos preferentes con formas gramaticales no aceptadas	0	19,8%	10,0%	15,8%
Términos preferentes con formas de singular y plural no aceptadas	0	4,7%	2,2%	40,5%
Términos preferentes que carecen de relación jerárquica	0	0%	1,4%	1,1%
Términos preferentes con dos o más términos genéricos	0	53,8%	14,9%	0%
Términos preferentes con un solo término específico	0	13,5%	3,5%	10,3%

Para el indicador (b) los tesauros Spines y el de Unesco tienen un comportamiento aceptable, mientras que el Decs está fuera de los límites del valor de referencia. De todas formas, los tres tesauros poseen términos con demasiadas palabras por término preferente.

El primer indicador (a) revela que cerca del 10% de los términos preferentes representan más de un concepto, cuando el valor de referencia es 0. Esta característica es propia de otros vocabularios, como las listas de epígrafes, pero no del tesauro.

Para el indicador (b) los tesauros Spines y el de Unesco tienen un comportamiento aceptable, mientras que el Decs está fuera de los límites del valor de referencia. De todas formas los tres tesauros poseen términos con demasiadas palabras por término preferente.

En cuanto a los términos preferentes con formas gramaticales no aceptadas (indicador c) los tres vocabularios transgreden la norma, siendo Spines el que más se acerca al valor de referencia.

En el indicador (d) Spines es el que más se acerca al ideal, presentando un 2,2% de formas no aceptadas. Un dato llamativo para este indicador es el tesoro de UNESCO con un 40,5% de formas no aceptadas en español, pero no en inglés, idioma en que sí respeta las normas.

En el indicador (e) DeCS cumple con la norma, mientras que en los otros vocabularios la transgresión representa un porcentaje bajo.

En el indicador (f) el tesoro de la UNESCO cumple con la norma, mientras que el Decs posee demasiados términos preferentes con dos o más términos genéricos (más del 50%).

En el último indicador presentado (g) si bien los tres vocabularios presentan transgresiones, Spines es el que más se acerca al valor de referencia.

En la Figura 1 se puede observar el comportamiento de los vocabularios con respecto de cada uno de los seis indicadores que se expresan como porcentaje.

Analizando en forma integral los 7 indicadores, el tesoro que mejor comportamiento presenta es el tesoro Spines ya que solo el valor del indicador f supera escasamente el 10%.

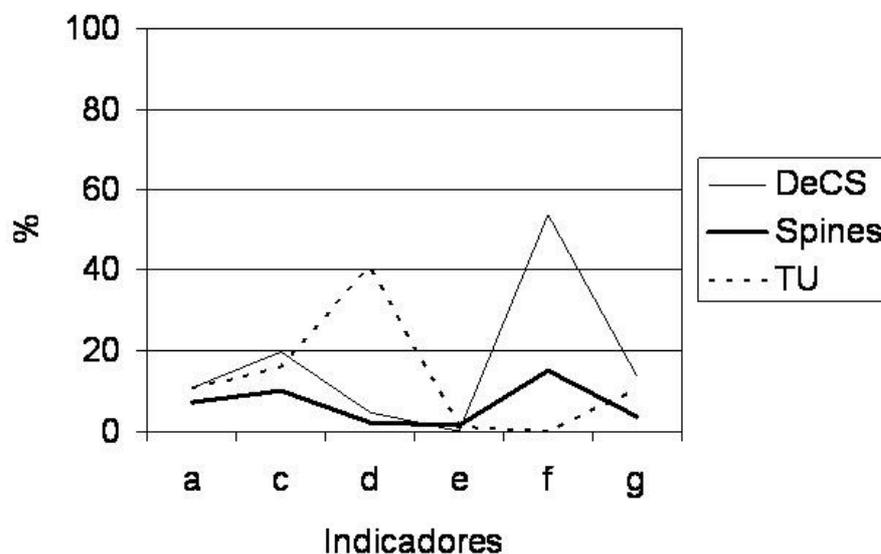


Figura 1. Comportamiento de tres tesauros respecto a seis indicadores

Conclusiones

Se proponen aquí siete indicadores, aplicables a la etapa de diseño y desarrollo del tesauro. Los indicadores propuestos en el presente trabajo han sido especialmente concebidos para la evaluación previa a la utilización del tesauro por parte de los futuros indizadores y usuarios del sistema de información. Estos indicadores cumplen con los criterios de Orr (1973) en cuanto a ser informativos, válidos, adecuados y comparables. Respecto a su nivel de confiabilidad, si bien han sido aplicados a tres tesauros distintos con resultados equivalentes, sería necesario reiterar su aplicación en otras experiencias desarrolladas por otras personas, sobre otros tesauros y durante un cierto tiempo, para poder afirmar que son verdaderamente confiables. En cuanto a la aplicabilidad, el indicador b (media de palabras por término preferente) resulta el más práctico de todos, ya que puede incorporarse fácilmente a cualquier software de gestión de tesauros. Si bien los indicadores restantes pueden demandar mayor intervención humana, el esfuerzo requerido es razonable, particularmente si se tiene en cuenta que el objetivo final es asegurar la calidad del tesauro.

Es importante destacar que tanto los requisitos contra los cuales se deben comparar las características de los tesauros, así como las pautas para evaluar la calidad de cualquier objeto, proceso o sistema, están especificados en normas nacionales e internacionales,

ampliamente conocidas, aceptadas y fáciles de consultar. De modo que la tarea de establecer indicadores y evaluar la calidad de los tesauros debe basarse en dichos requisitos documentados.

Agradecimientos

Proyecto PID H004 acreditado por la Universidad Nacional de La Plata.

Bibliografía

- Aitchison, J.; Gilchrist, A.; Bowen, D. 2000. Thesaurus construction and use: a practical manual. 4th ed. London; Aslib, p. 47-80.
- Alvaro Bermejo, C.; Villagra Rubio, A.; Sorli Rojo A. 1989. Desarrollo de lenguajes documentales formalizados en lengua española: II. Evaluación de los tesauros disponibles en lengua española. Revista Española de Documentación Científica. Vol. 12 (4), 283-305.
- Bireme: Centro Latinoamericano y del Caribe en Ciencias de la Salud. 2008. Descriptores en ciencias de la salud [en línea]. São Paulo; El Centro. <http://decs.bvs.br/E/homepagee.htm> [Consulta 21 Nov 2009].
- BSI Group. 2005. Structured vocabularies for information retrieval: guide. Part 2: thesauri. London; The Group. (BS 8723-2:2005).
- Chung, Y. M.; Lee, J. Y. 2004. Optimization of some factors affecting the performance of query expansion. Information Processing and Management. Vol. 40 (6), p. 891-917.
- Claverdon, C. W.; Mills, J. 2005. The testing of index language devices. En: Chan, L. M.; Richmond P. A.; Svenonius, E., eds. Theory of subject analysis: a sourcebook. Littleton: Libraries Unlimited, p. 221-246.
- Commission des Communautés Européennes. 1976. Définition des caractéristiques essentielles des thesauri: Rapport final. Bruxelles; Bureau Marcel van Dijk. Vol. 1.
- Gil Leiva, I. 2008 Manual de indización: teoría y práctica. Gijón; Trea, p. 213-217.
- Gil Urdiciain, B. 1998. Evaluación semántica y estructural de tesauros. Revista General de Información y Documentación. Vol. 8 (2), 193-199.

- Instituto Argentino de Normalización. 2007. Información y documentación: indicadores de desempeño de bibliotecas. Buenos Aires: El Instituto.
- International Organization for Standardization. 1986. Documentation: guidelines for the establishment and development of monolingual thesauri. 2nd ed. Geneva: The Organization. (ISO 2788:1986).
- International Organization for Standardization. 2000. Quality management systems: fundamentals and vocabulary. 2nd ed. Geneva; The Organization. Geneva: The Organization. (ISO 9000:2000).
- Lancaster, F. W. 2002. Control del vocabulario para la recuperación de información. 2a ed. Valencia; Università de Valencia, p. 31-37, 55-62.
- Mills, J. 2004. Faceted classification and logical division in information retrieval. Library Trends. Vol. 52 (3), 541-570.
- National Information Standards Organization. 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies [en línea]. Bethesda MD; NISO Press. (ANSI/NISO Z39.19:2005). <http://www.niso.org/standards/resources/Z39-19-2005.pdf> [Consulta 28 Nov 2009].
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. 1988. Tesauro Spines [en línea]. París; La Organización. http://thes.cindoc.csic.es/index_SPIN_esp.html [Consulta 17 Nov 2009].
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. 2007. Tesauro de la Unesco. 15ª ed. París; La Organización. <http://databases.unesco.org/thessp/> [Consulta 17 Nov 2009].
- Orr, R.H. 1973. Measuring the goodness of library services: a general framework for considering quantitative measures. Journal of Documentation. Vol. 29 (3), 315-332.
- Owens, L. A.; Cochrane, P. A. 2004. Thesaurus evaluation. En: Roe, S. K.; Thomas, A.R., eds. The thesaurus: review, renaissance, and revision. New York: Haworth Information Press, p. 87-102.
- Ranganathan, S. R. 1967. Prolegomena to library classification. (3rd ed.). New York; Asia Publishing House. Chapter EC, p. 145.

- Spiteri, L. A. 1998. simplified model for facet analysis. Canadian journal of information and library science. Vol. 23 (1), 1-30.
- Vickery, B. C. 1960. Faceted classification: a guide to the construction and use of special schemes. London: Aslib.