

UN LENGUAJE PARA LA ORGANIZACIÓN DEL CONOCIMIENTO

BIBLIOTECOLÓGICO EN AMÉRICA LATINA.

Catalina Naumis Peña
Universidad Nacional Autónoma de México.
Correo electrónico:
naumis@servidor.unam.mx

Maria Texia Iglesias Maturana
Universidad Tecnológica Metropolitana de Santiago de Chile.
Correo electrónico:
tiglesia@omega.utm.cl

RESUMEN:

Se presenta la metodología que se ha seguido en la construcción del Tesauro Latinoamericano en Bibliotecología y Ciencia de la Información, con el propósito de aportar aquellos elementos sistemáticos para la construcción de lenguajes capaces de comunicar el conocimiento, tomando en cuenta las diferencias que existen entre las variantes dialectales del español hablado en Latinoamérica. Esta ponencia también se propone dar a conocer ésta fuente de información para la Bibliotecología y la Ciencia de la Información. También se incluyen aquí los antecedentes y las estrategias de trabajo seguidas para culminar este trabajo de alcance internacional.

PALABRAS CLAVE: TESAURO, BIBLIOTECOLOGÍA, CIENCIA DE LA INFORMACIÓN

ABSTRACT:

The methodology followed in the building of the Thesaurus in Library and Information Science is hereby presented, with the purpose to contribute with systematic elements to the construction of languages capable of communicate knowledge taking into account the differences which exist among the spanish dialects spoken in Latin America. This paper also intends to disseminate this source of information among the professionals of library and information science. The antecedents and strategies followed to accomplish this work of international scope, are also included.

KEYWORDS: THESAURUS, LIBRARY SCIENCE, INFORMATION SCIENCE

INTRODUCCIÓN

La construcción de un tesauro requiere el empleo de métodos en los que prevalezca un alto grado de validación empírica de todos y cada uno de los elementos que lo componen, debido a que su estructura y aparato terminológico constantemente están siendo sometidos a pruebas de eficiencia mediante la práctica y la sistematización de la búsqueda y recuperación de la información. Estos métodos no son privativos para la construcción de tesauros, sino que también forman parte de los métodos científicos que por lo general utilizan casi todas las ciencias que pretendan la comprobación de una hipótesis y una aproximación a la verdad, sobre un proceso o fenómeno determinado, es por ello que para la construcción de tesauros se requiere de un fuerte componente teórico y disciplinario que sustente la estructura del conocimiento representada, así como de mecanismos de confirmación empírica de sus descriptores que corroboren su eficacia técnica en la práctica de la recuperación de información, aspectos éstos complejos que sólo pueden lograrse por medio de un método científico.

En una primera etapa del desarrollo de una disciplina, los conceptos del lenguaje ordinario pueden resultar suficientes para representar una especialidad, luego surgen los conceptos abstractos y la terminología técnica que los nombra. Desde el punto de vista de Hempel¹ "Para lograr teorías de gran precisión, amplio alcance y alta confirmación empírica, la ciencia debe producir, por lo tanto, en sus diferentes ramas, sistemas comprensivos de conceptos especiales, referidos por términos técnicos." Tanto en el proceso de adquisición de nuevos conocimientos, como en la revisión de lo sabido y conocido se estructuran los conceptos e integran nuevos términos para nominarlos y enriquecer el lenguaje formal de las disciplinas científicas, se aportan elementos terminológicos para la construcción de los lenguajes documentales y el posterior acceso a los nuevos conceptos incorporados al conocimiento.

Sin embargo, para los fines de este trabajo y la investigación, que sobre lenguaje se realiza para la organización del conocimiento bibliotecológico latinoamericano se utilizó una definición más pragmática de lenguaje especializado explicada en forma clara por Lera² "Desde el punto de vista lingüístico una terminología no se presenta a primera vista como un conjunto de nociones, sino como un conjunto de expresiones que sirven para denominar en una lengua natural las nociones que forman un área de conocimientos muy tematizadas," tratando con ello de no dejar fuera de la bibliotecología los textos para los no especialistas, que se producen al interior de la disciplina para tipos de usuarios como pueden ser los estudiantes de licenciatura de la misma.

Los términos especializados y técnicos utilizados por la Bibliotecología en español están muy representados en diccionarios, cuya diferencia señala, Rey "[...] son los niveles de uso, no las áreas jerarquizadas, son las colocaciones más frecuentes y más normales, no la sintagmática terminológica, y es el valor de empleo en el discurso, no la articulación conceptual, lo que prevalece y ha de prevalecer en esta óptica descriptiva."³

Mientras que el tesoro es un lenguaje concebido como instrumento de comunicación mediante los elementos siguientes: la representación de la estructura conceptual de un campo temático; la aclaración del sentido en el cual se debe utilizar el término; la eliminación de los equívocos mediante el direccionamiento de la equivalencia univalente clara entre los términos; las referencias recíprocas entre términos relacionados, es decir, el tesoro nos muestra el campo de aplicación de un término y su posición en la disciplina. Al respecto Sager⁴ expresa "[...] es posible crear un diccionario o un glosario sobre la base de un tesoro terminológico, pero nunca al revés, porque el diccionario alfabético no contiene la información relacional necesaria para abarcar el tema", es decir, el tesoro constituye por sí solo un sistema terminológico relacional orientado a ser un sistema de búsqueda de información enmarcado dentro de una disciplina genérica o específica.

La normalización del lenguaje de una disciplina para una región del mismo idioma, pero con variaciones en expresiones para nominar los mismos conceptos allana el camino para la creación de condiciones objetivas favorables a la cooperación entre países y a la transferencia de información. La formación de una zona hemisférica de libre comercio con la participación de todos los países del continente americano ha comenzado con el Tratado de Libre Comercio entre México y Estados Unidos de Norteamérica y el Mercosur, entre Argentina, Brasil, Paraguay y Uruguay y deberá continuar con la formación de un Mercado Común que llevará consigo la total integración económica, política y social de las Américas. Cuando esto ocurra la América hispano-hablante deberá estar preparada para defender su lengua y junto con ella, los valores culturales e identidad, así como mantener un desarrollo científico propio. La creación de lenguajes de búsqueda y recuperación de información, en castellano, constituye una pequeña aportación en este sentido.

El conocimiento científico y el desarrollo tecnológico determinan las condiciones y formas del dominio del hombre sobre la naturaleza, como consecuencia una sociedad bien informada sobre los avances en estas materias tendrá mayores posibilidades de crecimiento económico y social. Existe información en línea, centros de información, bibliotecas y bases de datos en esferas muy específicas del conocimiento; sin embargo se dedica tiempo cada vez más prolongado para localizar información

relevante, cuando la carencia más fuerte para el investigador de hoy en día es justamente el tiempo. Parece contradictoria esta afirmación si se considera la disposición de recursos bibliográficos tan poderosos por la cantidad de registros y entradas para acceder información, cuya función facilita la recuperación y posterior consulta de las fuentes primarias.⁵

La contradicción consiste no tanto en la disponibilidad de recursos y tecnologías, sino en la calidad del procesamiento y sistematización de la información; es decir, que la presencia de una indexación temática profunda, basada en lenguajes documentales apropiados para el acceso, constituye en última instancia la clave de la eficiencia de los sistemas de recuperación.

El éxito de un servicio de información y referencia estará determinado, en buena medida, por la calidad de organización y sistematización de los documentos y fuentes de información que lo integran. En este sentido, un lenguaje normalizado de búsqueda informativa asegura una concepción clara de las temáticas tratadas en un sistema de información, tanto al analista como al usuario del mismo.

El Centro Universitario de Investigaciones Bibliotecológicas perteneciente a la Universidad Nacional Autónoma de México apoya un Proyecto Regional Cooperativo: INFOBILA (Información y Bibliotecología Latinoamericana), base de datos bibliográficos que busca integrar la producción científica latinoamericana que se genera sobre Bibliotecología y Ciencia de la Información, como uno de los medios para desarrollar la infraestructura que le permita cumplir con sus funciones de investigación, docencia y difusión de la especialidad. Participan en el Proyecto varios países de América Latina constituidos en red quienes aportan los registros bibliográficos de los documentos producidos en cada uno de ellos.

Uno de los aspectos prioritarios en el marco de INFOBILA ha sido la normalización del lenguaje utilizado en el análisis de los documentos y, como consecuencia, el desarrollo de un vocabulario controlado, tipo descriptor, que responda a dos objetivos básicos: 1) servir de tesoro oficial para la indexación de los documentos ingresados en la Base de Datos INFOBILA, y 2) lograr una normalización de la terminología de la especialidad utilizada en los países de la región. Dada la importancia que tiene para la disciplina el desarrollo de este proyecto, el objetivo del presente trabajo es presentar la metodología utilizada, y las estrategias de trabajo que se han seguido en la construcción de este tesoro.

ANTECEDENTES Y PRINCIPALES FUENTES UTILIZADAS EN LA CONSTRUCCIÓN DEL TESAURO DE BIBLIOTECOLOGÍA Y CIENCIA DE LA INFORMACIÓN

La responsabilidad para la elaboración de este Tesoro Latinoamericano en Bibliotecología y Ciencia de la Información fue asignada a un grupo de trabajo con experiencia en el desarrollo de tesauros y que pertenece a diferentes instituciones, a saber: Escuela de Bibliotecología de la Universidad Tecnológica Metropolitana de Chile, Centro Universitario de Investigaciones Bibliotecológicas de la UNAM, Biblioteca Nacional de Venezuela y Facultad de Comunicación y Lenguaje de la Pontificia Universidad Javeriana de Colombia.

Las disciplinas bibliotecológica y de la información se han desarrollado en los últimos veinte años, con mucha fuerza; este crecimiento ha estado caracterizado por el uso generalizado de computadoras. La penetración de la tecnología y de cómputo en estas disciplinas ha traído consigo cambios en el lenguaje formal de la misma, sin embargo este cambio acelerado en terminología no se ha reflejado en una alta producción de tesauros o vocabularios especializados.

Los resultados de la búsqueda de instrumentos lingüísticos en América Latina para la indexación y recuperación de información en Bibliotecología y Ciencia de la Información fueron poco exitosos.

En 1980, auspiciado por la Federación Internacional de Documentación y publicado por el

Instituto Colombiano para el Fomento de la Educación Superior en Colombia, se publicó un tesoro en el área de Documentación e Información.⁶ En el año de 1996 la terminología presentada era insuficiente para su utilización.

En 1982, el Centro Universitario de Investigaciones Bibliotecológicas de México se publica un Vocabulario controlado en bibliotecología, ciencia de la información y temas afines,⁷ destinado para su utilización en la base de datos INFOBILA. Este vocabulario fue un buen comienzo para la organización temática de la información y ayudó a mejorar la calidad de la indización y recuperación. Fue realizado a partir de un estudio terminológico comparativo sobre los términos que utilizaban los indizadores de la base de datos INFOBILA. Una vez publicada la mencionada obra, se prosiguió con la construcción de una lista de términos candidatos proveniente del análisis de los documentos y las consultas realizadas a la base. La cantidad de documentos analizados y el mismo uso de la base es una fuente inagotable de terminología para normalizar. La necesidad de mantenerlo actualizado y mejorar su estructura como tesoro fue uno de los alicientes para proseguir con el trabajo.

El Instituto Brasileiro de Informação em Ciência e Tecnologia publicó en 1989 el *Tesouro de ciência da informação*,⁸ para sustentar la indización temática de su base de datos de la disciplina en idioma portugués, por lo tanto podía servir como punto de apoyo para la elaboración de un tesoro en idioma español en lo referente a su estructura temática y jerárquica, pero no podía ser aplicado para su uso en la base de datos INFOBILA.

Ahora bien en Chile y Colombia, las experiencias a tomar en cuenta fueron el resultado de trabajos de titulación, posteriormente desarrollados y ampliados por las profesoras guías. Por lo tanto, el propósito primario de estos tesoros fue de orden académico, y no de difusión para un uso generalizado.

Mientras tanto en Venezuela se integraba una lista de términos elaborada por los indizadores del CEDINBI, cuya terminología surge a medida que se indizan cantidades importantes de documentos sobre Bibliotecología y Ciencia de la Información, que ingresan a la Biblioteca Nacional de Venezuela

En vista de los mencionados antecedentes y considerando la insuficiencia de estos lenguajes, así como la necesidad de unificar estas experiencias en un lenguaje que satisfaga las necesidades de la región en la disciplina, se consideró la conjunción del trabajo de personas diferentes que realizaron su labor en distintos países, con diversos propósitos y para instituciones, ya sea de servicio, enseñanza o de investigación, logrando con ello constituir un grupo de trabajo dinámico, responsable de la elaboración del Tesoro en Bibliotecología y Ciencia de la Información.

Una vez constituido el grupo, la estrategia de trabajo establecida consistió en la organización de talleres de trabajo de dos semanas, en México en 1996, en Chile en 1997 y en Venezuela en 1998.

METODOLOGÍA UTILIZADA EN LA CONSTRUCCIÓN DEL TESAURO

Para elaborar el tesoro se utilizó la combinación de dos métodos: el deductivo, a partir de los temas principales en que el comité estableció la armonización de las jerarquías y subdividió la disciplina; y el método inductivo o empírico para la recolección de los términos provenientes de diversas fuentes.

Para la preparación del tesoro se contó con las fuentes mencionadas en los antecedentes y con listas de palabras clave extraídas de los títulos y resúmenes, con su correspondiente frecuencia de aparición en la base de datos INFOBILA. Se trabajó con el sistema tradicional denominado "comité de expertos", en el que las decisiones y discusiones en torno a la conveniencia de la incorporación de cada uno de los descriptores al lenguaje normalizado fueron producto de un análisis extenso, con los especialistas de cada una de las áreas temáticas comprendidas en el tesoro.

Las etapas que se siguieron en la construcción del tesoro y a las que se debió adaptar el comité, estructurado a distancia, fueron las siguientes: definición del campo de trabajo, recolección, normalización, organización de los términos, evaluación del tesoro.

El campo de trabajo fue definido a partir de las políticas de ingreso de documentos en la base de datos INFOBILA. Una vez recolectados los términos candidatos cada término fue analizado mediante el examen del concepto que servía de base para la designación lingüística. Se buscó un punto de vista único para la clasificación de los términos a partir de la percepción de los usos dialectales de cada uno.

El acuerdo de los miembros del Comité fue sostenido, una alta proporción de veces, por las significaciones validadas en las fuentes de consulta de la especialidad y por la consulta a expertos en las diferentes temáticas incluidas en la disciplina bibliotecológica.

Fueron decisiones de los miembros del Comité: el análisis y clasificación de los términos para ser ingresados o no en el sistema terminológico, las estrategias para representarlos y la interpretación de las normas para su utilización en el tesoro; la aceptación de sinónimos no ingresados como entradas principales, pero referidos a los aceptados; las categorías de los términos con su ubicación jerárquica; el análisis de las propiedades para establecer relaciones conceptuales y la redacción de notas de alcance para explicar el sentido en el que deben ser utilizados los descriptores.

Para las evaluaciones primarias se contó con la participación de las instituciones que apoyan el proyecto, sin descartar la posibilidad de evaluaciones más profundas, una vez liberada la primera versión para su uso en Latinoamérica.

ETAPAS DE DESARROLLO DEL TESAURO

Es importante mencionar las etapas de desarrollo del tesoro, por su perspectiva, en cuanto que se trata de un proyecto internacional que puede servir como antecedente para experiencias similares. El trabajo estuvo dividido en tres partes:

ETAPA DE RECOLECCIÓN

- Elaboración de una metodología para el desarrollo del tesoro.
- Identificación y evaluación de las áreas del conocimiento que serían incluidas en el tesoro para la base de datos INFOBILA.
- Establecimiento de criterios para seleccionar los términos que presentan diferencias de concepto, según la opinión de los miembros del comité.
- Elaboración de una lista alfabética de la letra A a la letra Z, integrada por los diferentes términos validados o no, integrantes de los tesauros aportados por los países participantes, agregando los descriptores utilizados en la base de datos Infobila. En los casos de sinonimia se escogió el término más generalizado en el ámbito latinoamericano y se incluyeron los otros como términos no aceptados.
- Selección del software que sería utilizado y determinación de la conveniencia de unificar el uso del mismo para todos los países integrantes del proyecto.
- Distribución de la tarea de cada país participante, para la evaluación del inventario terminológico.

ETAPA CONCEPTUAL:

- Se observaron los resultados de aplicación del inventario terminológico con los siguientes

experimentos : base de datos INFOBILA en el CUIB, colección de la Biblioteca Nacional de Venezuela y ejercicios con los alumnos de la especialidad en la Escuela de Bibliotecología y Documentación de Chile.

- Se discutió la inclusión de los términos sugeridos por cada institución participante.
- Se redujeron al mínimo las partes de indeterminación del significado de los términos mediante la clarificación de las funciones expresivas de cada uno, a través de una red de relaciones.
 - determinaciones directas notas de alcance
 relaciones jerárquicas
 - determinaciones indirectas términos relacionados
 términos equivalentes

ETAPA DE EVALUACIÓN

- Se revisaron e intercambiaron opiniones sobre cada uno de los términos candidatos y las listas elaboradas por las instituciones participantes. Se analizaron uno por uno, para decidir su aprobación o rechazo e incorporación al cuerpo del tesoro, ya sea en calidad de preferentes o no usados
- Se establecieron las relaciones correspondientes de los descriptores validados y efectuaron las modificaciones pertinentes en el corpus alfabético del tesoro, justificando además los descriptores no aprobados.
- Se construyeron las designaciones de los términos genéricos superiores que encabezarían las jerarquías de las categorías principales y se aprobó la estructura temática siguiente:

ANÁLISIS Y SISTEMATIZACIÓN DE LA INFORMACIÓN
FUENTES Y DOCUMENTOS
ADMINISTRACIÓN
SERVICIOS
FUNDAMENTOS TEÓRICOS
TECNOLOGÍA DE LA COMUNICACIÓN Y LA INFORMACIÓN
TRANSFERENCIA DE LA INFORMACIÓN
PROFESIÓN

CONSIDERACIONES FINALES

Más allá de la utilidad que este tesoro pueda tener para la indización y recuperación de documentos de la especialidad, subyacen otros objetivos muy importantes: lograr una normalización en el lenguaje utilizado por los profesionales de información en América Latina; ser utilizado como una fuente para el análisis de la indización, búsqueda y recuperación de información. Si estos objetivos se logran, los resultados obtenidos constituirán un gran aporte para el intercambio de experiencias profesionales, académicas y de investigación en la región.

La detección terminológica de palabras significativas de las disciplinas es una actividad muy socorrida por programas de cómputo dedicados a ello, pero el análisis de las significaciones de los términos y la conveniencia de su uso para la representación y recuperación de información es una tarea humana, compuesta de procesos interpretativos y creativos de difícil dilucidación

La creación de un tesoro para una disciplina o para una base de datos no es un fin en sí, sino un medio para procesar la información y luego accederla con rapidez y eficiencia. Se utiliza un tesoro para dar respuesta a las búsquedas temáticas de los usuarios, sobre un fondo documental especializado en algún área del conocimiento o en una base de datos que responda al interés de un grupo de disciplinas. Esto no implica que los descriptores que integran el cuerpo alfabético de un tesoro sean todos términos especializados, pero aun en los casos de términos del lenguaje común

deben ser acotados a los usos en la disciplina. En una investigación sobre el porcentaje de términos que constituyen un sublenguaje científico, Haas⁹ observó que cerca del 25% de las expresiones de un resumen son términos propios de la disciplina e incluso en un diccionario de computación, sólo entre el 24% y el 66% designa conceptos especializados del tema.

La construcción de tesauros se apoya principalmente en la terminología, al utilizar sus métodos para crear y consolidar las bases terminológicas que se requieren en el desarrollo de los conceptos especializados unívocos, en la teoría del conocimiento para la ramificación de las especialidades, en la lingüística para la sintaxis de las expresiones y en la semiótica para determinar el requisito que cada elemento de indización represente o se dirija únicamente hacia un concepto y provoque en los usuarios el estímulo esperado. Los especialistas en información organizarán la estructura del tesoro y aportarán los puntos de vista de los indizadores y usuarios del sistema de información.

El Tesoro Latinoamericano en Bibliotecología y Ciencia de la Información ha sido concluido con 1180 términos aceptados y 600 rechazados e incluye una amplia red de relaciones para contextualizar los términos del lenguaje controlado y reducir al mínimo las partes de indeterminación del significado de los descriptores.

¹ Hempel, Carl G. "Fundamentos de la formación de conceptos en ciencia empírica". – Madrid: Alianza Editorial, 1988. – p. 38

² Lerat, Pierre "Las lenguas especializadas". – Barcelona: Ed. Ariel, 1997. –p.17

³ Rey, Alan "La terminologie dans un dictionnaire général de la langue française le Grand Robert". – TermNet News, (14), 1995. p. 6.

⁴ Sager Juan C. "Curso práctico sobre el procesamiento de la terminología". – Madrid : Fundación Sánchez Ruy Pérez, 1993 – p 283

⁵ Cousins, Shirley Anne Enhancing subject access to OPACS : controlled vocabulary vs natural language, 1992, p. 292.

⁶ Rozo Ríos, Dora Susana y María Eugenia Garzón Herrán "Tesoro en Documentación e Información". Bogotá: Federación Internacional de Documentación, Comisión Latinoamericana : Instituto Colombiano para el Fomento de la Educación Superior, 1980.

⁷ Peniche de Sánchez Macgregor, Surya "Vocabulario controlado en bibliotecología, ciencia de la información y temas afines". México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 1992.

⁸ Gusmao, Heloísa Ríos Tesoro de ciencia da informacao : consideraciones teóricas e comentarios. "Transinformacao", 1(3), 1989, p. 43-48.

⁹ Haas Stephanie W. y Shaoyi He Toward the automatic identification of sublanguage vocabulary. – "Information Processing Management", 29, (6), 1993, p. 723