



EL COLEGIO DE MÉXICO, A. C.

BIBLIOTECA DANIEL COSÍO VILLEGAS

**EL LENGUAJE DE LA INDIZACIÓN Y EL LENGUAJE DE LA BÚSQUEDA POR
TEMAS: IMPLICACIONES DE SU ENCUENTRO EN LA RECUPERACIÓN
TEMÁTICA**

**(ANÁLISIS EN EL CATÁLOGO DE LA BIBLIOTECA DANIEL COSÍO
VILLEGAS)**

Tesis presentada por

ADRIANA SUÁREZ SÁNCHEZ

Para optar por el grado de

MAESTRA EN BIBLIOTECOLOGÍA

Director de tesis

MTRO. REYNALDO D. FIGUEROA SERVÍN

Lectores

**MTRA. LOURDES GUERRERO ANDRADE
(EL COLEGIO DE MÉXICO)**

**DR. JOHN WRIGHT
(BRIGHAM YOUNG UNIVERSITY)**

MÉXICO, D.F.

ABRIL DE 2011

Gracias:

A la vida por su gran generosidad conmigo.

A mis hermanas: Erika, Rosa Velia y Clara. ¿Qué sería de mí sin sus locuras y elocuencias?

A mis padres, Filiberto y Rosa Velia, por su apoyo.

A Aramiz por su amor y paciencia.

A Camila, donde quiera que esté, Sofi-lu y Musho.

A mis queridos amigos, pocos pero efectivos.

A mis compañeras de generación: Daniela, Alexa y Cristina.

A mis profesores y tutores de la maestría por sus enseñanzas y consejos.

A El Colegio de México y la Biblioteca Daniel Cosío Villegas.

ÍNDICE

Introducción	1
1. Fundamentos teórico-conceptuales relacionados con la recuperación temática en bibliotecas	7
1.1. La indización en las bibliotecas	8
1.1.1. Concepto	8
1.1.2. Aspectos epistemológicos de la indización de documentos	10
1.1.3. El proceso de indización de documentos	16
1.1.3.1. El análisis documental	16
1.1.3.2. La asignación de temas	20
1.2. La búsqueda temática	21
1.2.1. Concepto	21
1.2.2. La búsqueda por temas en catálogos de bibliotecas	25
1.3. El lenguaje en la recuperación por materias	28
1.3.1. El lenguaje de la indización	28
1.3.2. El lenguaje de la búsqueda temática	31
1.3.3. El lenguaje de indización frente al lenguaje de búsqueda de información	32
1.4. Conclusión	38
2. Encuadre institucional	40
2.1. El Colegio de México y la Biblioteca Daniel Cosío Villegas	41
2.2. La indización en la BDCV	43
2.3. La búsqueda por temas en el catálogo de la BDCV	45
2.4. El lenguaje de la indización y de la búsqueda por temas en la BDCV	47
2.5. Conclusión	50
3. Fundamentos metodológicos	51
3.1. Instrumentos de análisis	52
3.1.1. Transacciones de búsquedas por temas	52
3.1.1.1. Justificación	52
3.1.1.2. Acopio de datos	52
3.1.1.3. Preparación de datos	54
3.1.1.4. Análisis de datos	54
3.1.1.5. Presentación de resultados	56

3.1.2. Cuestionario	56
3.1.2.1. Justificación	56
3.1.2.2. Diseño	57
3.1.2.3. Acopio de datos	58
3.1.2.4. Análisis de datos	59
3.1.2.5. Presentación de resultados	59
3.2. Conclusión	59
4. Resultados obtenidos mediante las transacciones de búsquedas por temas	61
4.1. La petición de búsqueda	62
4.1.1. Problemas frecuentes	62
4.1.1.1. Idioma	62
4.1.1.2. Naturaleza de la colección	64
4.1.1.3. Búsquedas mal dirigidas en el campo temas	66
4.1.1.4. Estado de la formulación	66
4.1.2. Estructura	69
4.1.2.1. Palabras simples	70
4.1.2.2. Frases	71
4.1.2.3. Dos o más palabras enlazadas	72
4.1.2.4. Palabras compuestas	73
4.1.2.5. Siglas	73
4.1.2.6. Impacto de la estructura en la recuperación de documentos	73
4.1.3. Características	75
4.1.3.1. Búsquedas por tópico	76
4.1.3.2. Búsquedas personales	77
4.1.3.3. Búsquedas geográficas	79
4.2. Coincidencia de la petición de búsqueda con la indización	80
4.2.1. Coincidencia nula	80
4.2.2. Coincidencia parcial	82
4.2.3. Coincidencia mediante autoridad temática	83
4.2.4. Coincidencia completa	83
4.3. Conclusión	84
5. Resultados obtenidos mediante el cuestionario sobre la búsqueda por temas	85

5.1. Uso de la búsqueda	86
5.2. Opinión de los usuarios sobre la búsqueda	89
5.3. Formulación de la búsqueda	91
5.4. Petición de búsqueda	95
5.5. Conocimientos del usuario sobre el lenguaje de indización	96
5.6. Recuperación en la búsqueda	98
5.7. Sugerencias de los usuarios para mejorar la búsqueda	101
5.8. Conclusión	101
Conclusiones generales	103
Bibliografía	106
Anexos	114

Introducción

Esta investigación tiene sus raíces en mi deseo de indagar sobre la función de los lenguajes documentales en las bibliotecas. Desde mi ingreso a la maestría, ya entonces con la firme convicción de realizar una tesis en este campo y, posteriormente, tras los cursos de Organización del conocimiento y Organización de la Información, surgió en mí un mayor interés por los lenguajes documentales en éstas áreas. Así, poco a poco, el tema de la tesis fue modificándose hasta concretarse en este trabajo que trata los problemas producidos por el encuentro del lenguaje del usuario y el lenguaje de indización en la recuperación temática.

Llevar a cabo un estudio de esta naturaleza resulta importante, ya que las bibliotecas, en general, dedican gran cantidad de recursos (monetarios, tecnológicos y humanos) a la clasificación temática; no obstante, no se tiene una idea precisa sobre cómo esta actividad está siendo útil y cómo es percibida por los usuarios. Actualmente, en muchas bibliotecas la indización o catalogación temática es una actividad primordial dirigida a la organización y recuperación de los materiales. En tal sentido, en México hay bibliotecas que cuentan con índices de calidad; sin embargo, contar con buenos índices no resulta suficiente, es necesario saber si este trabajo se transforma en beneficios para los usuarios. Al respecto, Luis Villen Rueda señala que:

Los estudios sobre la recuperación de información en los catálogos de bibliotecas forman parte de la evaluación en bibliotecas. Tarea de especial importancia para medir la calidad de los servicios y en lo que se refiere a procesos técnicos mediante esta actividad se pueden medir los alcances, limitaciones y ventajas o desventajas que la actividad ofrece a los usuarios [...] la única forma de saber si como profesión, se está trabajando en la línea correcta es examinar los servicios que se ofrecen a los usuarios. Este es un aspecto de la investigación en el área de la documentación y, específicamente, de la recuperación de información que corre en paralelo a la evolución de la profesión y de los medios tecnológicos para desarrollarla.¹

En México se desconoce si todo el trabajo invertido en la indización está siendo útil y menos aún se cuenta con investigaciones más específicas como la que se plantea esta tesis. Sobre la búsqueda por temas la literatura dice que es de las opciones más complicadas. Al respecto, Karen Antell y Jie Huang² han atribuido su carácter complejo a la dificultad que le implica al usuario la selección de términos adecuados para la localización de documentos. También sobre el tema, Mari Carmen Marcos señala que “la recuperación de información relevante en los OPAC resulta difícil para los usuarios

¹ Luis Villen Rueda. “Indización y recuperación por materias en los OPACS de las bibliotecas españolas: ¿dos décadas de evaluación?” en *El profesional de la información* (Barcelona: Swets & Zeitlinger Ibérica, vol.15, núm. 2, 2006), pp. 87-98.

² Karen Antell y Jie Huang. “Subject searching success: transaction logs, patron perceptions and implications for library instruction” en *Reference Users Service* (Chicago: American Library Association, vol. 48, núm. 1, 2008), pp. 68-76.

cuando se trata de búsquedas por materias ya que en la mayoría de los casos los usuarios desconocen los descriptores o términos empleados por los catalogadores para indizar los documentos.”³

En nuestro país, se percibe que un gran número de la bibliografía se centra en el proceso de indización, dejando de lado los estudios de evaluación de la actividad en función de su utilidad los usuarios. Los primeros trabajos que intentaron vincular la indización con la recuperación por temas fueron llevados a cabo en Estados Unidos y estuvieron dirigidos a probar la efectividad de los encabezamientos de la biblioteca del congreso (LCSH), mediante la comparación de los términos que el usuario selecciona y los de la lista de encabezamientos.

Un trabajo pionero, fue realizado en 1941 por Eugene Wall y Jean Barnes.⁴ Se llevó a cabo en la National Agricultural Library e indagó sobre la coincidencia entre vocabularios especializados de diversas disciplinas (Agricultura, Medicina, Biología) y los encabezamientos de la Biblioteca del Congreso. Su objetivo era analizar si los encabezamientos podían dar a cabida a todos los temas de las disciplinas. Los resultados fueron que sólo el 6.1% de los términos de la disciplina encontraban correspondencia en los encabezamientos.

Otro trabajo importante se hizo en 1980 y fue realizado por Tillie Krieger⁵ quien comparó los términos que el usuario usa para describir documentos y los aceptados por LCSH. El experimento consistió en solicitar a un grupo de usuarios la asignación de descriptores para representar una serie de documentos. Una vez que ellos asignaron los términos, se comparó su coincidencia con LCSH y se encontró que sólo el 13.9% de los descriptores seleccionados por los usuarios coincidía con la lista de encabezamientos.

En la década de los noventa Marilyn A. Lester⁶ hizo un estudio en el que analizó la correspondencia entre las palabras de búsquedas de los usuarios y la indización, basaba en LCSH. Sus resultados muestran que el 40% de los vocablos introducidos por los usuarios se enlazan con un término de los índices temáticos. Este

³Mari Carmen Marcos. *Mejoras en la consulta y presentación de los resultados en catálogos de bibliotecas* (Barcelona: Sección Científica de Ciencias de la Documentación., Departamento de Ciencias Políticas y Sociales, 2003. (DE, 13 de septiembre, 2009: www.mcmarcos.com/pdf/2003_mejoras-vigo.pdf)).

⁴Eugene Wall y Jean Barnes. *Intersystem compatibility and convertibility of subject vocabularies*. (Philadelphia: Auerbach Corp, 1969.).

⁵Tillie Krieger. *Instructor influences versus text influences in the selection of subject descriptors by undergraduate students* (Tesis doctoral: Universidad de Illinois, Urbana-Champaign, 1981.).

⁶Ann Marilyn Lester. *Coincidence of user vocabulary and Library of Congress subject headings: experiments to improve subject access in academic library online catalogs* (Urbana, Ill. University of Illinois at Urbana Champaign, 1989).

estudio hizo evidente la falta de coincidencia entre el lenguaje de búsqueda y el lenguaje de indización.

En el periodo del 2000 a la fecha hay varios trabajos⁷ que relacionan la indización y la búsqueda por temas; no obstante, siguen siendo investigaciones norteamericanas y europeas.

En nuestro la investigación al respecto, como ya se mencionó, se encuentra en un estado incipiente. Y la posibilidad de que en años próximos se realice es incierta, dado que en México sólo unas cuantas bibliotecas (académicas y de investigación) cuentan con recursos para realizarla. Pero, aunque el panorama se vislumbra complicado, emprender estos trabajos es indispensable para saber si las actividades de las bibliotecas están favoreciendo a los usuarios.

Ante este panorama, la presente tesis busca indagar sobre un aspecto específico de la recuperación por temas en catálogos de bibliotecas. **El objetivo general de mi investigación es: determinar las implicaciones o problemas del encuentro entre el lenguaje del usuario y el lenguaje de indización en la recuperación temática.**

Además del objetivo general, se persiguen algunos objetivos específicos:

- a) Establecer cómo impactan estos problemas en la recuperación por temas.
- b) Señalar en qué porcentaje de búsquedas se presentan los problemas.
- c) Esbozar, en la medida de lo posible, soluciones a los problemas presentados.

Las preguntas de investigación que guían el trabajo son:

- a) ¿Qué tipo de palabras introduce el usuario en las búsquedas temáticas?
- b) ¿El tipo de palabra que introduce en su búsqueda influye en la recuperación de información?
- c) ¿En el sistema es posible el enlace entre la petición de búsqueda y la indización?
- d) ¿El factor “coincidencia entre el término introducido por el usuario y el término de indización” influye en la recuperación por temas?
- e) De las búsquedas por tema, ¿qué cantidad presenta problemas relacionados con el encuentro entre el lenguaje de indización y el lenguaje del usuario?
- f) ¿Qué puede hacer una biblioteca para que haya mayor vinculación entre las peticiones del usuario, expresadas en lengua natural, y la indización, realizada en lenguaje controlado?

⁷ Karen Antell y Jie Huang. “Subject searching success...”. *Op. cit.*

Para cumplir los objetivos planteados y dar respuesta a las preguntas de investigación, **se emplean dos instrumentos de análisis: transacciones de búsquedas por temas y cuestionario**. Cabe señalar que el foco de la investigación está centrado en el usuario ya que en torno a él giran las actividades que se desarrollan en las bibliotecas.

El objeto de estudio de la investigación es el catálogo en línea de la Biblioteca Daniel Cosío Villegas (BDCV) de El Colegio de México (COLMEX). Por estar centrada sólo en una biblioteca, la tesis bien puede enmarcarse en un estudio de caso; no obstante, por ser de las primeras investigaciones sobre el tema es un antecedente para trabajos posteriores en las que podrían realizarse estudios comparativos. También es importante porque es de los primeros estudios en el contexto mexicano que usan las transacciones de búsqueda como instrumento de análisis y propone un cuestionario sobre aspectos concretos de la búsqueda temática.

Una limitación a la que se enfrentó la tesis fue que, en función de los tiempos establecidos para ser completada, no se hizo una estratificación de usuarios (investigadores, estudiantes de doctorado, maestría y licenciatura, becarios y usuarios externos) sino que se agruparon todos como usuarios finales. Un seguimiento de cada tipo de usuario hubiera sido una variable extra, pues la literatura indica que los usuarios con mayor conocimiento terminológico de su disciplina de estudio y con mayores habilidades informativas logran enlazar mejor sus peticiones con la indización hecha por la biblioteca.

La tesis, finalmente, se presenta en cinco capítulos:

1. Marco teórico: en él se delimitan conceptualmente todos los elementos relacionados con la indización, la búsqueda por temas, el lenguaje de la indización y el lenguaje de la búsqueda por temas. También se establecen relaciones entre los distintos elementos involucrados en la recuperación por materias.
2. Encuadre institucional: presenta el contexto donde se llevó a cabo la investigación.
3. Marco metodológico: ofrece información detallada sobre los instrumentos de análisis.
4. Resultados de las transacciones de búsquedas por temas: muestra los datos obtenidos mediante el instrumento.

5. Resultados obtenidos mediante el cuestionario: ofrece información derivada del segundo instrumento.

En cada capítulo se añade un apartado denominado “Conclusión” en el que se presentan los aspectos más importantes del apartado. Aunado es esto, hay una “Conclusión general” que integra los resultados parciales de los distintos capítulos.

**Capítulo1. Fundamentos teórico-conceptuales
relacionados con la recuperación temática en
bibliotecas**

1.1. La indización en las bibliotecas

1.1.1. Concepto

No es posible hablar de una definición concertada de indización, pues durante el desarrollo de la bibliotecología a lo largo de muchos siglos, y en épocas recientes de las llamadas ciencias de la información, ha sido definida de diversos modos por los autores y las corrientes de la disciplina o según las exigencias de las investigaciones que se han valido del vocablo.

En la década de los setentas, la *Encyclopedia of Library and Information Science* definió la indización como “el proceso de crear entradas en un índice”⁸ y la caracterizó como una tarea basada en cuatro actividades: revisión del documento, análisis del contenido, etiquetado discreto de los ítems en la colección e incorporación del identificador para la precisa localización de los materiales. Según esta definición, la única finalidad de la indización era posibilitar la localización de materiales en las colecciones. De cierto modo, su conceptualización se corresponde con el contexto de las bibliotecas en esa época, caracterizado por colecciones pequeñas, procesos manuales y catálogos en tarjetas.

Luego de la explosión informativa y la automatización de bibliotecas, ocurridas a principios de los años ochentas, el concepto de indización se modifica. La *International Encyclopedia of Library and Information Science*, dice que es: “El análisis del contenido de un documento (libro, panfleto o ítem audiovisual o legible por máquina) o una colección de documentos y la traducción del resultado del análisis en términos que permitan la localización y recuperación de la información.”⁹ Además de mencionar el análisis de contenido como punto de partida de la actividad, señala que, luego de definir el asunto de un libro, se hace un proceso de “traducción” en el cual el contenido debe ser trasladado a términos. A diferencia del concepto anterior, esta definición establece dos finalidades: localización y recuperación de los materiales. Aunque dice que la recuperación es una finalidad de la indización, en otras secciones de la obra, al hablar específicamente sobre los propósitos de ésta, dice que: “el principal propósito [de la indización] es la rápida localización de ítems específicos de información”.¹⁰

⁸ Allen Kent, Harold Lancour y Jay Daily, eds. *Encyclopedia of library and information science* (Nueva York: Marcel Dekker, 1974), s. v. Indización.

⁹ John Feather y Paul Sturges. *International encyclopedia of information and library science*. (Nueva York: Routledge, 1997), s. v. Indización.

¹⁰ *Idem.*

Otro autor, J. M. Reitz, tanto en su obra *Dictionary for Library and Information Science* como en el *ODLIS: Online Dictionary for Library and Information Science*,¹¹ presenta dos definiciones. La primera de ellas, ligada a la construcción de índices, es:

El proceso de compilar uno o varios índices de una sola publicación, como una monografía o trabajo de referencia en varios volúmenes, o la adición de otras entradas para nuevos documentos que anteriormente ya tenían una entrada en el índice, pueden ser obras con un formato de publicación determinada (por ejemplo, periódicos), trabajos de un formato literario específico (biografía, reseñas de libros, etc.) o la literatura de un ámbito académico, disciplina, o grupo de disciplinas.¹²

La segunda, más vinculada al análisis documental, es la siguiente:

Un proceso donde un profesional entrenado, indizador, lee o escanea el texto de un documento para determinar su contenido, entonces selecciona los encabezamientos apropiados (nombres, lugares, temas) para facilitar la recuperación la referencias cruzadas son hechas para los sinónimos y las entradas son arregladas en la secuencia deseada (alfabética, numérica, clasificada, etc.). En una entrada del índice, los descriptores del contenido son usualmente seleccionados de una lista de términos preferidos (vocabulario controlado) desarrollado a través del tiempo por el servicio de indización. La indización puede ser pre-coordinada o post-coordinada.¹³

Reitz retoma un viejo problema conceptual de las ciencias de la información, pues es común en la literatura de la disciplina encontrar ambigüedad entre la indización entendida como la actividad de compilar índices y la indización como la actividad en la que se asignan términos útiles para la recuperación del documento.¹⁴ De las definiciones enunciadas por Reitz, la más útil a este trabajo es la segunda, en tanto que en ella define la indización como una actividad en la cual un indizador determina el contenido del texto y elige palabras útiles para su recuperación. También es valiosa porque incorpora dos aspectos vitales: la creación de referencias cruzadas y los tipos de indización: precoordinada o poscoordinada.

En lengua española, una obra que incorpora la palabra indización es la de J. López Yepes, quien la define como una:

Técnica de Análisis Documental que describe y representa el contenido de las fuentes de información mediante un número limitado de conceptos extraídos del texto de los documentos (palabras clave) o de vocabularios controlados (clasificaciones, listas de materia, tesauros), que van a permitir el control y la recuperación de información de un conjunto documental dado.¹⁵

¹¹ Joan M. Reitz. *ODLIS: Online dictionary for library and information science* (Westport, CT: Libraries Unlimited, 2004-2010. (DE, 20 de Agosto, 2010: <http://lu.com/odlis/>)), s. v. Indización.

¹² Joan M. Reitz. *Dictionary for library and information science* (Westport, CT: Libraries Unlimited, 2004), s. v. Indización.

¹³ *Idem*.

¹⁴ En algunas obras este problema conceptual se resuelve usando términos distintos para los procesos. Así, la indexación es el proceso de construcción de índices y la indización alude a la determinación temática o catalogación por temas.

¹⁵ José Lopez Yepes, ed. *Diccionario enciclopédico de ciencias de la documentación* (Madrid: Síntesis, 2004), s. v. Indización.

López Yepes, encuadra la indización dentro de una actividad superior - el Análisis Documental- y en su definición contempla características de la indización relacionadas con el lenguaje usado en ésta. A partir de esta característica, establece dos tipos de indización. La primera, representa el contenido del documento con palabras clave del mismo, mientras que la segunda incorpora herramientas externas a éste (clasificaciones, listas de encabezamientos y tesauros). Así, distingue entre indización controlada e indización libre.

Sobre la indización controlada, dice que es:

Una técnica de Análisis Documental, realizada en la fase de tratamiento de los sistemas de información, dirigida a representar el contenido de un documento mediante un número limitado de conceptos extraídos de vocabularios controlados o lenguajes documentales (clasificaciones, listas de materia, tesauros) para permitir su recuperación posterior.¹⁶

Y la indización libre, que bien puede ser vista como opuesta o complementaria de la indización controlada, es:

Una técnica del Análisis documental, realizada en la fase de tratamiento de los sistemas de información, dirigida a representar el contenido de un documento mediante un número limitado de conceptos extraídos del texto de los documentos (palabras clave) para permitir su recuperación posterior.¹⁷

Una vez que se han revisado varias definiciones de indización resulta necesario presentar la que se usará en este trabajo. Es importante mencionar que es una definición propia, adecuada a los fines y propósitos de la tesis. Por indización me refiero a:

Un proceso que consiste en analizar un documento con la intención de identificar partes de contenido que sirvan para representarlo mediante la asignación de términos- libres o controlados- a través de los cuales pueda ser recuperado en un índice. La indización se produce en dos niveles: análisis documental y asignación temática. En el primer nivel se analizan partes del documento para determinar de qué trata y se conceptualiza el contenido del documento en una descripción, mientras que en la segunda parte hay un proceso de traducción de la descripción mediante la asignación de términos de indización.

Finalmente, cabe señalar que la indización es un proceso complejo del que dependen muchas funciones de la biblioteca y es vital en la recuperación por temas en los catálogos.

1.1.2. Aspectos epistemológicos de la indización de documentos

Como se estableció en el apartado anterior, la indización busca posibilitar la localización y recuperación de documentos en un sistema de información. Para cumplir sus metas, se vale de representaciones temáticas de entidades documentales. En la vida

¹⁶ *Idem. s. v.* Indización controlada.

¹⁷ *Idem. s. v.* Indización libre.

cotidiana se representan entes reales (un árbol, una casa, un auto, etc.) con la finalidad de asignarles nombres con los que sea posible su comunicación y manipulación. En la representación documental se persigue los mismos fines, sólo que las entidades son materiales (libros, revistas, discos, folletos, documentos electrónicos, etc.) que se conceptualizan y simbolizan.

En bibliotecología, decidir cómo deben ser representadas temáticamente las entidades (documentos) es complicado, pues no hay maneras únicas para hacerlo. Se parte de que toda representación es artificial; en el sentido de que no es la realidad, sino un reflejo de ésta. Aunado a esto, toda representación es en cierto grado subjetiva, pues siempre está sujeta a la apreciación de quien la hace. Estas características serían observables si pidiésemos a varias personas que mediante un dibujo hagan la representación del referente “casa (objeto de la vida real)”. Es probable que los dibujos obtenidos sean distintos, ya que estarán influidos por el contexto de quien lo realiza. De modo que si bien puede haber representaciones temáticas de los documentos, es probable que éstas sean diferentes en tanto que cada persona, indizador en el caso de la representación documental, la realiza de acuerdo a ciertas normas y su universo contextual.

En la representación de entidades y la representación de documentos están presentes tres elementos que son tomados en cuenta por el representador. Éstos son, el referente: es una entidad que forma parte del mundo, puede ser un referente real (libro) o un referente imaginario (conocimiento); el concepto: una idea dentro de la estructura de conocimiento de una mente o varias y el símbolo: palabra o palabras que representan el referente. Las relaciones entre los elementos fueron establecidas en el triángulo semiótico propuesto por Ogden y Richards,¹⁸ quienes redujeron estas proposiciones al esquema siguiente:

¹⁸ Charles Kay Ogden y I.A. Richards. *El significado del significado: una investigación acerca de la influencia del lenguaje sobre el pensamiento y de la ciencia simbólica* (Buenos Aires: Paidós, 1954), pp. 339-340.

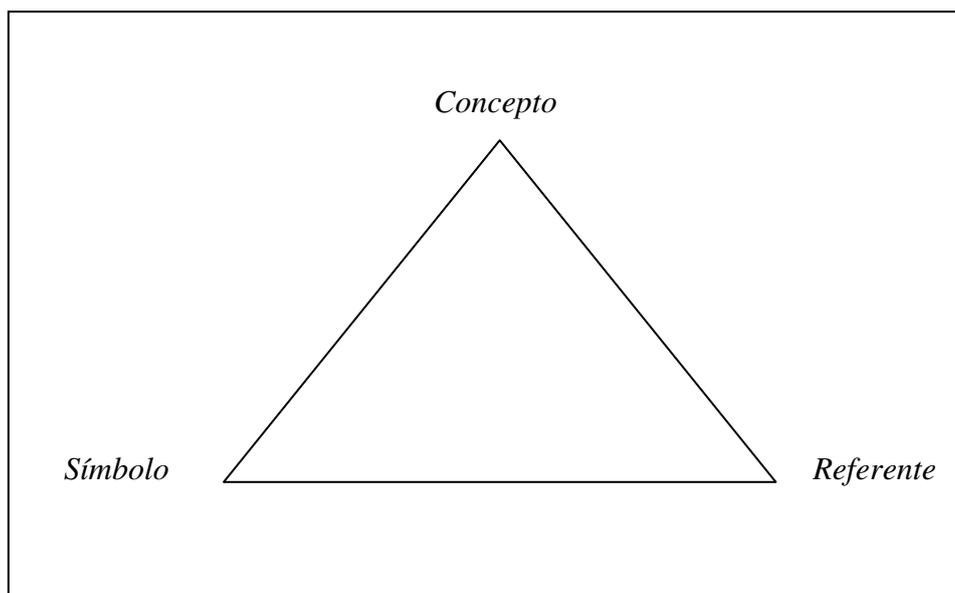


Fig. 1. Triángulo semiótico de Ogden y Richards

El inicio de la representación está en un referente al que el representador se enfrenta en la vida diaria. En la cotidianidad los referentes pueden ser reales o abstractos, en las ciencias documentales, los referentes son entidades reales, es decir, recursos documentales; sin embargo, su representación no se encuentra ligada al objeto en sí (libro, revista, disco compacto, etc.) sino al contenido albergado en éste. Así, para identificar los referentes de un documento el representador tiene que ir al *aboutness*,¹⁹ punto de partida para la representación. Mientras que en la realidad los referentes son únicos, esto es, una mesa no puede ser a la vez una silla, en las ciencias documentales los materiales pueden tener un referente o varios, es decir, tratar un tema o más de uno. Es por eso que los distintos referentes de un documento pueden dar lugar a múltiples representaciones.

Así como un documento puede tener varios referentes, un mismo referente puede dar lugar a variedad de conceptos, esto es, un documento puede ser conceptualizado de diverso modo, según el entramado cognitivo del representador. Por ejemplo, un libro sobre “Yihad”, cuyo título sea sólo este, puede ser conceptualizado de distintas maneras. Imaginemos que tal vez un indizador norteamericano haga una equiparación inmediata entre “Yihad” y “Terrorismo”, ya que su contexto cultural lo inclina a establecer esta relación, mientras que un indizador cuyo contexto histórico y cultural es diferente posiblemente esté más inclinado a asociar “Yihad” con los “Aspectos religiosos de la guerra” o “Islam y Estado”.

¹⁹ *Aboutness* es el asunto sobre el que trata un documento. En el siguiente apartado se le tratará ampliamente.

La conceptualización de un documento es una idea mental que el indizador forma, basándose en su contenido. Al respecto, Mario G. Barite señala que en toda conceptualización, es posible identificar dos caras: una interna y otra externa. La interna refiere, al “conjunto de enunciados que determinan al concepto, que se suceden desde el más general hasta el más específico y son necesarios para identificar al referente de modo inequívoco,”²⁰ mientras que la cara externa “corresponde a la palabra, el conjunto de palabras u otro tipo de símbolo que identificará y representará al concepto.”²¹ Esta cara externa o “fachada del concepto” es la que facilita y simplifica los procesos de comunicación entre dos o más personas.

En el plano de lo ideal la relación referente-concepto-símbolo debería ser univoca, empero, no es así, pues es común encontrar que un referente puede tener más de un concepto y varios símbolos mediante los que puede ser representado. Explico lo enunciado:

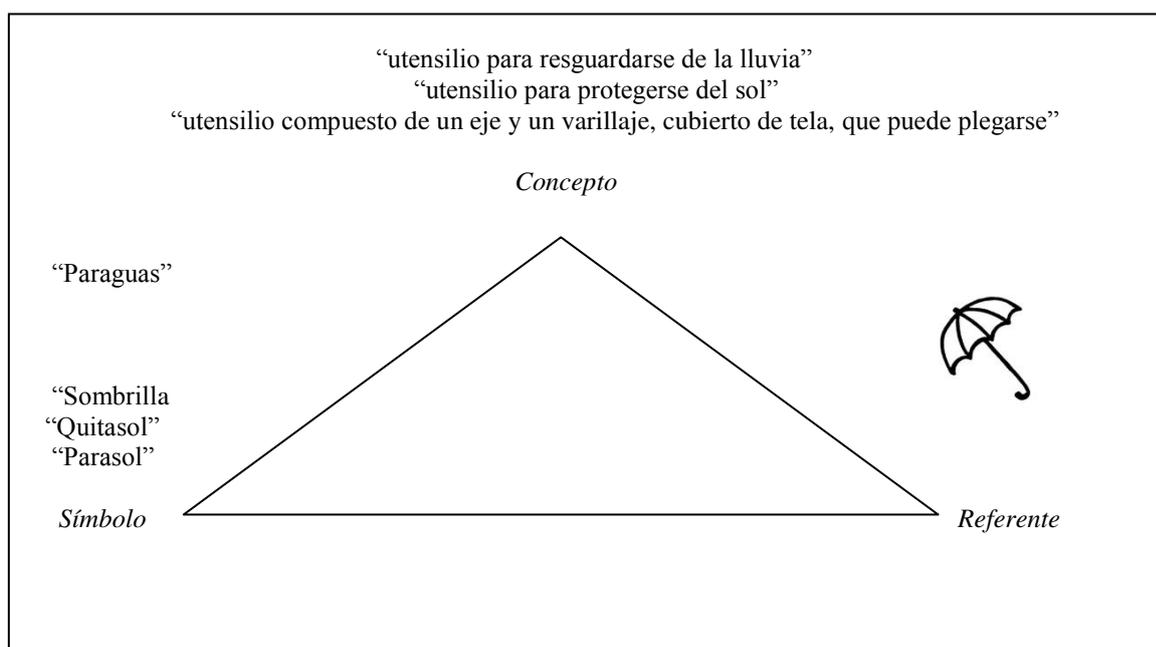


Fig. 2. Triángulo de la representación

Al inicio de la representación hay una entidad: el objeto. Luego, uno o varios conceptos: “utensilio para resguardarse de la lluvia”, “utensilio para protegerse del sol” o “utensilio compuesto de un eje y un varillaje, cubierto de tela, que puede plegarse”. Se percibe que un mismo referente puede ser conceptualizado de diversas maneras. El

²⁰ Mario Guido Barite Roqueta. “Los conceptos y su representación: una perspectiva terminológica para el tratamiento temático de la información” en *Scire* (Zaragoza, España: Universidad de Zaragoza núm. 6, vol. 2, 2000), pp. 31-53.

²¹ *Idem.*

tercer elemento del triángulo es el símbolo, concebido como “una expresión lingüística o simbólica, representación del objeto y de su concepto para fines de comunicación”²². Sobre este tercer elemento se observa que un mismo concepto puede ser representado por símbolos diferentes. Mientras que en los conceptos “utensilio para resguardarse de la lluvia” y “utensilio para protegerse del sol” los símbolos serían *paraguas* y *sombrilla*, respectivamente, el tercer concepto “utensilio compuesto de un eje y un varillaje, cubierto de tela, que puede plegarse” puede tener cualquiera de los dos. También se observa la presencia de varios símbolos (sinónimos) para el concepto “utensilio para protegerse del sol”, éstos son *sombrilla*, *quitasol* y *parasol*.

A partir del esquema, se concluye que la representación en la vida diaria y por extensión en la representación documental no mantienen una relación unívoca sino que se enfrentan a problemas de conceptualización y simbolización de los referentes. En la indización este hecho da lugar a que un mismo referente pueda ser representado de diversos modos.

La terminología es la disciplina que más se ha ocupado del control de los distintos símbolos para representar conceptos. A lo largo de muchos años, bajo la teoría tradicional, buscó consolidar una relación unívoca entre el referente, el concepto y el símbolo. El máximo exponente de esta corriente fue Eugene Wüster, quien postuló que a cada concepto sólo podía corresponder un término y viceversa, eliminando todas las posibles relaciones de homonimia, polisemia y sinonimia. Sin embargo, en años recientes surgió la teoría cognitiva que aceptó la sinonimia, homonimia y polisemia como fenómenos de la lengua natural, pero enfatizó que con fines de control terminológico deben controlarse.

Aunque el problema de diversos símbolos para representar un concepto y diversos conceptos vinculados a un referente afecta directamente a la terminología, su alcance se extiende a todas aquellas disciplinas que en alguna de sus actividades se valen del uso de términos. La bibliotecología es una de ellas, dado que una de sus tareas fundamentales es representar y recuperar entidades bibliográficas mediante el empleo de términos.

La representación de entidades en bibliotecología se encuentra estrechamente relacionada con la indización ya que su objetivo es auxiliar en la organización de la información con miras a hacer los documentos localizables y recuperables para los

²² *Idem.*

usuarios.²³ La representación como parte de los procesos de la biblioteca puede ser observada en la catalogación; en la catalogación descriptiva sirve para representar datos tales como autor, título, editorial etc., y en la catalogación temática consiste en “la determinación del tema tratado en una pieza bibliográfica, o en los trabajos que esta contiene, con el fin de clasificarla y determinar el encabezamiento o los encabezamientos adecuados de materia por los que va a estar representada en un catálogo.”²⁴

La representación temática, materializada en la indización es una operación en la cual un catalogador extrae información relevante del contenido de una fuente, recurso impreso o electrónico, y la simboliza con ayuda de un lenguaje de indización. Sobre la función de la representación en la indización de documentos, José Lopez Yepes señala que “toda representación es limitada y se centra en los aspectos de la realidad que interesan a las necesidades y motivaciones del representador.”²⁵ Al hablar de un representador nos referimos a un individuo o una institución que puede ser la biblioteca. En bibliotecología la representación busca ser lo más objetivo posible, de ahí que para realizar la actividad se cuente con herramientas, normas y estándares; no obstante, hay ocasiones en las que un mismo referente es representado de manera distinta.

El lenguaje natural es el instrumento idóneo en la representación de entidades, empero, no siempre es el empleado. Tal es el caso de la bibliotecología, donde muchas veces las entidades son representadas mediante lenguajes artificiales. Éstos tienen su origen en el lenguaje natural, pero se rigen por una sintaxis y semántica particular establecida en función de las necesidades de la disciplina. La representación de los referentes en las disciplinas que usan lenguajes especializados -lenguajes controlados en el caso de la bibliotecología- queda determinada por un grupo de símbolos posibles y

²³ En este trabajo los usuarios de las bibliotecas son aquellas personas que hacen uso de los servicios que ésta ofrece. Sin embargo, dado que esta investigación estudia los catálogos de bibliotecas como sistemas de recuperación de información, la definición que se toma de usuario se encuentra inserta en este contexto. Así el usuario es aquella persona que hace uso del catálogo. A él están dirigidas todas las acciones del sistema, pues su objetivo es organizar información para su recuperación, tornándose un agente transmisor de información entre los recursos y el usuario. También desde la perspectiva de los sistemas de recuperación de información, hay distintos tipos de usuarios: usuarios intermedios y usuarios finales. Los usuarios intermedios son quienes construyen el sistema, lo prueban, alimentan o hacen uso de éste para llevar a cabo su trabajo en tanto que los usuarios finales son aquellos a los que está dirigido el sistema como producto. Sintetizo, entonces, que en este trabajo el usuario es aquella persona, usuario final, que hace uso del catálogo como sistema de recuperación de información.

²⁴ Heartsill Young. *Glosario ALA de bibliotecología y ciencias de la información* (Madrid: Díaz de Santos, 1988), s. v. Descripción temática.

²⁵ José Lopez Yepes, ed. *Diccionario enciclopédico de... Op.cit. s.v. Representación.*

aceptados. Este hecho dificulta la tarea del representador, pues, por un lado, tiene que asociar el referente a un concepto y representarlo con un símbolo en su lenguaje natural y, por otro, tiene que hacer una traducción o equivalencia de ese símbolo en lenguaje natural al lenguaje de especialidad.

1.1.3. El proceso de indización de documentos

1.1.3.1. El análisis documental

El proceso de indización se constituye de dos etapas: análisis documental y asignación de temas. El análisis documental es “la parte de la indización o la catalogación que se encarga de determinar de qué trata un documento.”²⁶ Sobre cómo un catalogador lleva a cabo esta tarea hay poco escrito en español. Algunos autores en lengua inglesa como J.W. Hutchins y Alenka Šauperl conscientes del estado incipiente del tema han indagado un poco más. No obstante aún hay mucho por trabajar; el mismo Hutchins dice que: “hemos indagado muy poco sobre cómo los indizadores y clasificadores deciden cuál es la materia de un documento y cómo ellos deciden sobre qué trata éste”.

El análisis documental tiene como meta determinar el *aboutness* de los ítems. El concepto de *aboutness* fue introducido a principios de la década de los setentas por Fairthorne, empero, ha sido Hutchins quien más ha investigado al respecto. Para él es:

Una expresión que resume el contenido de un documento [...] el objeto de la indización (sea automatizada o no) se concibe como la provisión de una expresión o un conjunto de palabras clave que en conjunto representan un resumen del contenido de un documento.²⁷

Entre autores y clasificadores que escriben en lengua inglesa han surgido discusiones sobre si *aboutness* es la palabra adecuada para referirse al asunto o materia sobre el cual trata un documento o debiera usarse el término *subject matter*. En este sentido, Hutchins considera que debe preferirse el término *aboutness*, ya que “en indización una cosa debe ser nombrada de manera clara, en este sentido, la noción de materia es relativamente vaga, podría ser cualquier tema tratado en un documento, mientras que el *aboutness* refiere al asunto global o generalizado al que refiere el contenido de éste.”²⁸ Birger Hjørland apoya esta idea y señala que “el argumento para

²⁶ Luis Villen Rueda. *La indización y el acceso por materias en los catálogos de bibliotecas: el desafío de la recuperación de información* (Guadalajara, Jalisco, México: Universidad de Guadalajara, 2006), p. 41.

²⁷ J.W. Hutchins. “The concept of ‘aboutness’ in subject indexing” en *Aslib Proceedings* (Londres: ASLIB vol. 30, núm. 5, 1978), pp. 172-181.

²⁸ *Idem.*

introducir el concepto en ciencias de la información se debió a dificultades conceptuales que implica la palabra tema.”²⁹ En español, se ha adoptado el vocablo anglosajón y también se usa “de qué trata un documento” como su traducción más cercana. Aunque se usan como sinónimos las palabras tema, materia o asunto, de igual manera como sucede en inglés son vocablos vagos que no engloban el contenido de un material.

El proceso para determinar el *aboutness* de un documento está determinado por actividades realizadas por el catalogador como parte del análisis documental. Alenka Šauperl,³⁰ quien en 1999 hizo un estudio cualitativo sobre cómo un grupo de catalogadores expertos determinaban de qué tratan los documentos, concluyó que es una actividad constituida por varias etapas. En la primera el catalogador establece contacto con el documento a través de la lectura del título y poco a poco el contacto se va enriqueciendo mediante la revisión de la tabla de contenido, la lectura del prefacio, resúmenes, notas y, en algunos casos, incluso, la lectura de apartados completos del documento. Una vez que el catalogador ha hecho esta revisión está en posibilidad de saber de qué trata el material. Entonces, viene una segunda etapa que Šauperl llama “determinación de conceptos más importantes” en la cual el indizador en una palabra o palabras describe el documento.

Hutchins coincide con este proceso propuesto por Šauperl, pues dice que es:

Un proceso en el que el indizador debe realizar una serie de pasos para discriminar qué parte del libro es aquella de la cual se puede extraer el *aboutness*, una reducción del contenido a palabras que lo engloben y una etapa de adecuación de las palabras elegidas a un lenguaje controlado.³¹

Hasta ahora se ha hablado de un solo *aboutness*; sin embargo, es pertinente decir que no es único, es decir, un documento puede tener más de uno en virtud de quien lo determina. Siguiendo la teoría de Ingwersen,³² en cada recurso se puede hablar de varios *aboutness*: el *aboutness* del autor, el *aboutness* del indizador, el *aboutness* de la interrogación en el sistema y el *aboutness* del usuario.

²⁹ Birger Hjørland. *Information seeking and subject representation: an activity theoretical approach to information science* (Westport, Ct.: Greenwood Press, 1997), p. 63.

³⁰ Alenka Šauperl. *Subject determination during the cataloging process* (Tesis de doctorado: University of North of Carolina, 1999.).

³¹ J.W. Hutchins. “The concept of ‘aboutness’...”. *Op.cit.*

³² Peter Ingwersen. “Cognitive perspectives of document representation”. En *CoLIS 4: fourth international conference on conceptions of library and information science: emerging frameworks and methods* : (Seattle WA, 21-25 July 2002) (Seattle, WA.: Universidad de Washintong), pp. 285- 300.

El *aboutness* del autor refiere al contenido del documento y bien podría ser la representación del documento a partir de las palabras presentes en éste. Se observa, en la tabla de contenido de los materiales y su estructura responde al orden dado al documento según las estructuras de organización textual: introducción, apartado teórico, sección metodológica, etc.

Un segundo *aboutness* es el del indizador, quien trata de construirlo a partir de la pregunta ¿de qué trata el documento? Ingwersen señala que “no es necesariamente la información del texto o la imagen original, sino ésta analizada y mezclada con la interpretación del indizador y los esquemas del conocimiento.”³³ En los modernos sistemas de recuperación de información (SRI) el *aboutness* del indizador suele ser el enfoque dominante; no obstante, en años más recientes ha empezado a surgir un nuevo paradigma de indización centrada en el usuario.³⁴ Si antes se asumía que el *aboutness* del indizador y el del usuario eran iguales, en este nuevo paradigma se postula que pueden ser diferentes. Por lo tanto, la indización debe incluir tanto el *aboutness* del indizador como otros *aboutness* que sean relevantes. Otra autora que apoya este paradigma es Karen Sparck Jones quien acuñó el vocablo *user related aboutness* que propone la integración del usuario al establecer el *aboutness*. La integración del usuario se haría tomando en cuenta el contenido del documento y su vinculación con cómo este será usado. También Hutchins considera que muchas veces la indización se centra demasiado en el *aboutness* del indizador y al respecto propone que:

La definición del *aboutness* no debe ser completamente neutral: el indizador debe hacer algunos supuestos acerca de qué aspectos del documento pueden ser de interés para los usuarios, él debería tomar en cuenta los conocimientos generales o promedio de los usuarios.³⁵

El tercer *aboutness* es el de la petición o consulta (*query*). Se observa justo en el momento en que un usuario introduce una petición en el catálogo. Su manifestación es a través de una palabra a partir de la cual espera recuperar información. Esta palabra busca un enlace entre la indización y la petición de búsqueda. Si la consulta obtiene respuesta, el sistema recupera documentos y si no la tiene, la búsqueda obtiene cero resultados. Este *aboutness* es producto de la transformación que el usuario hace de su necesidad en palabras que introduce al sistema.

³³ Peter Ingwersen. *Information retrieval interaction* (Los Ángeles, CA.: Taylor Graham, 1992), p. 51.

³⁴ Uno de los más importantes defensores de este paradigma es Peter Ingwersen quien establece la teoría cognitiva de recuperación de la información. Para ampliar el tema véase Spink, Amanda y Charles Cole, eds. *New directions in cognitive information retrieval* (Dordrecht, Netherlands: Springer, 2005.).

³⁵J. W. Hutchins. “The concept of ‘aboutness’ ...”. *Op. cit.*

El último *aboutness* es el del usuario y “está asociado con procesos que conducen a la generación de una consulta, como parte del modelo de recuperación de información”.³⁶ Una vez que el sistema ha generado una serie de resultados con base en la petición del usuario, éste verifica si los documentos obtenidos son lo que esperaba obtener a partir del término que introdujo en el sistema. El usuario confronta los resultados y decide si éstos sirven o no a su necesidad de información. Su *aboutness* está determinado por la relevancia que ante su problema o necesidad tienen los documentos recuperados por el sistema.

Una vez que se han expuesto los distintos *aboutness* que puede tener un documento: autor, indizador, interrogación y usuario, es posible percibir que cada uno de los sujetos que tienen establecen relación con el documento tiene una idea particular respecto al contenido de los documentos. Este hecho interviene en la exploración por temas, pues un autor puede considerar que su obra trata sobre un tema, el catalogador, experto en la actividad, y con base en una estructura de organización del conocimiento, determinaría uno o varios temas y, en el otro sentido, un usuario establece un *aboutness* de su necesidad que materializa en un *aboutness* de interrogación al sistema. Hay, por tanto, cuatro posibles *aboutness* de un recurso que pueden ser esquematizados del siguiente modo:

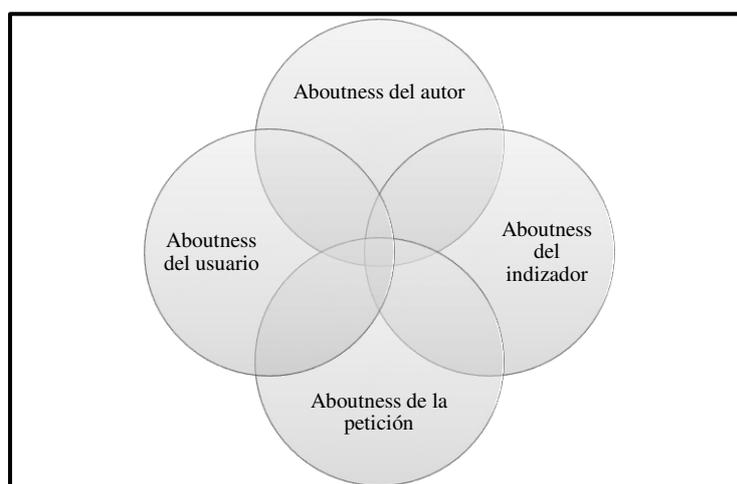


Fig. 3. Los *aboutness* de un documento

Aunque los distintos *aboutness* parecieran estar completamente desligados, entre ellos hay cierto grado de traslape, en tanto que existe un “conocimiento público”³⁷ sobre

³⁶ Peter Ingwersen. *Information retrieval interaction... Op.cit.* p. 55.

³⁷ Es un concepto usado por Vickery (Vickery, B.C. “Knowledge representation a brief review” en *Journal of Documentation* (Londres: Association of Special Libraries and Information Bureaux vol. 42, núm. 3, 1986), pp. 145-159.), quien dice que la representación de información se basa en un conocimiento

la relación de un referente, su concepto y los símbolos que lo representan. Así, el objetivo de la indización debe ser, por un lado, tomar en cuenta los posibles aboutness y, por otro, preocuparse por encontrar este punto de traslape a partir del cual la catalogación por temas sea amplia, integradora y posibilite un alto grado de recuperación.

Por tal motivo, no debemos hablar de representación sino de poli-representación, cuya finalidad es:

Mejorar la interacción entre el usuario y el sistema de recuperación [...] Debido a la individualidad de las estructuras de conocimiento de cada usuario, los tipos de información integrados en un sistema de recuperación deben ser representados de diferentes maneras mediante la aplicación de diferentes métodos de representación.³⁸

Según la poli-representación, hay diferentes maneras de percibir un objeto, es decir, focalizando distintas propiedades mediante las cuales puede ser indizado. Una gran ventaja de la poli-representación es que persigue una comunicación más efectiva, que posibilite la integración de los distintos aboutness. Tradicionalmente la indización tomaba como válido un solo *aboutness* (el del indizador); sin embargo, ya se empieza a considerar la integración de otros,³⁹ lo que será benéfico para la disciplina pues permitirá una indización más amplia de los documentos y repercutirá en mejores resultados en la obtención de registros.

1.1.3.2. La asignación de temas

Una vez que el indizador ha determinado el *aboutness* de un documento, la asignación temática completa el proceso. Ésta “consiste en asignar los temas que servirán para la localización y recuperación de acuerdo a un sistema de indización.”⁴⁰

En la asignación de temas el *aboutness* es traducido a términos libres o contenidos en listas de encabezamientos o tesauros. La asignación temática puede hacerse en lenguaje libre o controlado, cuando se trata de lenguaje libre se establecen temas sin control de vocabulario, pero cuando se trata de indización controlada es más

público, es decir, una especie de consenso sobre las entidades, esto, ejemplificado, sería que para la mayoría de la gente un objeto (silla) es una *silla* (símbolo) en tanto que tiene una base sostenida por patas y un respaldo y no un *sillón*, dado que el sillón es más grande y acojinado y tampoco sería un *banco*, pues el banco no tiene respaldo. También es un concepto usado por John Ziman en (Ziman, John M. *Public knowledge: an essay concerning the social dimension of science* (Cambridge, Mass.: Cambridge University, 1968.)).

³⁸Peter Ingwersen. “Cognitive Perspectives of ...”. *Op.cit.*

³⁹ En este trabajo, al referirme a *aboutness* asumo la idea de que en la determinación de éste se integran los distintos *aboutness* y coincido con quienes proponen que esta definición sea empleada en la indización.

⁴⁰ Peter Ingwersen. *Information retrieval interaction...* *Op.cit.* p. 40.

complicado ya que en este se presentan problemas de ajuste terminológico. Llega un momento crucial en el que el catalogador debe transformar un grupo de palabras en un encabezamiento o descriptor. Sobre el asunto, Hutchins indica que entre los factores de error, omisión o inconsistencia en el análisis y posterior representación temática de un documento está el hecho de que “el bibliotecario se preocupa más en ajustarse a un sistema de clasificación o de indización que en desarrollar un análisis documental que le permita, antes que nada, comprender el contenido expuesto en una obra.”⁴¹

En la asignación temática el indizador sigue tres principios: precisión, utilidad y claridad. Bajo el principio de precisión la entidad debe quedar fielmente representada; el principio de utilidad señala que el vocabulario utilizado en las descripciones y accesos debe estar en concordancia con el de los usuarios y bajo el concepto de claridad se entiende que los términos asignados como temas deben ser comprensibles. Es muy importante que el indizador haga bien su trabajo al seleccionar los encabezamientos o descriptores que dará al libro, pues de no hacerlo el usuario recuperará materiales que no son útiles a su necesidad de información.

1.2. La búsqueda temática

1.2. 1. Concepto

Definir la expresión no resulta sencillo, pues la literatura sobre la disciplina no la integra. Para rastrear su definición es necesario ir a un término más amplio: la búsqueda de información, ya que la opción por temas es una de las varias disponibles cuando se busca información.

La búsqueda de información, según el glosario de la ALA, es el “examen sistemático de los registros de un fichero o catálogo con el fin de encontrar registros o datos específicos.”⁴² Otra definición la considera “El hecho de que un usuario, un trabajador de la información en lugar del usuario o un sistema automatizado activado por el usuario o un intermediario haga una investigación sistemática para obtener datos o información.”⁴³ De las definiciones anteriores se infiere que es el escrutinio de registros en busca de información. La información a buscar puede ser precisa, como cuando se intenta localizar un libro del que se tiene el título exacto, pero también puede

⁴¹ J.W Hutchins. “On the problem of ‘aboutness’ in document analysis” en *Journal of Informatics* (Statesboro, GA. : International Academy for Information Management vol.1, núm.1, 1977), pp. 17-35.

⁴² Heartsill Young. *Glosario ALA de... Op.cit. s.v.* Búsqueda.

⁴³ John Feather y Paul Sturges. *International encyclopedia of...Op.cit. s.v.* Búsqueda.

ser una búsqueda exploratoria en la que se introduce una palabra clave y a partir de ésta se hace una exploración del sistema. Hay algunas obras que diferencian entre buscar con clara intención y una simple exploración. Sobre el asunto, Reitz dice que se debe recurrir a la palabra búsqueda cuando se escudriña con el objeto de encontrar algo ya mentalizado. Para él una búsqueda, en el sentido estricto de la palabra, sería: “un esfuerzo sistemático de parte de un usuario de la biblioteca o un bibliotecario para localizar información deseada por medios manuales o electrónicos, exitoso o no, este “esfuerzo” se opone a una revisión en la colección de la biblioteca sin una clara intención en la mente.”⁴⁴

La búsqueda de información es una actividad motivada y con finalidad, o sea, se da a partir de una necesidad y su finalidad es dar respuesta a ésta. J. López Yepes toma en cuenta esta característica cuando dice que: “es el procedimiento mediante el cual podemos obtener el conjunto de informaciones y documentos necesarios para resolver cualquier problema de investigación o información que se nos plantee.”⁴⁵

A grandes rasgos, es una actividad de solución de problemas en la que intervienen varios factores o elementos: un buscador de información, un problema o necesidad informativa, un área de conocimiento, un sistema de búsqueda y un conjunto de resultados, todos y cada uno de los cuales influyen en los demás y, por tanto, en el resultado final del proceso. Cabe destacar que el punto de partida de una búsqueda y por extensión de una búsqueda temática es una necesidad de información⁴⁶ que se desea satisfacer.

Una necesidad de información, tomando como idónea la definición de Belkin, es “un estado anómalo del conocimiento” que pasa por distintos lapsos hasta materializarse. Las etapas por las que atraviesa ésta son: una necesidad de información no manifestada; una descripción interna de la necesidad; una necesidad transformada en

⁴⁴ Joan M Reitz. *Dictionary for library... Op.cit. s.v.* Búsqueda.

⁴⁵ José Lopez Yepes, ed. *Diccionario enciclopédico de... Op.cit. s. v.* Búsqueda.

⁴⁶ El término necesidad de información ha sido conceptualizado de diversas formas: como una carencia o vacío, según Dervin (Dervin, B. “Information as user construct the relevance of perceived information needs to synthesis and interpretation” en Ward, S.A y Reed J. L., eds. *Knowledge structures and use. Implications for synthesis and interpretation* (Filadelfia: Temple University, 1983), pp. 153-184.); una necesidad visceral, menciona Taylor (Taylor, R.S. “The process of asking questions” en *American documentation* (Washington, D.C.: American Documentation Institute, vol. 36, núm. 3, 1962), pp. 391-397.); una anomalía en el estado del conocimiento dice Belkin (Belkin, N. J. “Anomalous state of Knowledge as a basis for information retrieval” en *Canadian journal of information science* (Toronto: Canadian Association for Information Science, vol. 5, núm. 3, 1980), pp. 133-143.); una situación problemática señala Wersing, quien analiza las necesidades de información en ciencias sociales, (Wersing, G. “The problematic situation as a basic concept of information science in the framework of the social sciences” en *New trends in informatics and its terminology* (Moscú: VINITI, 1979). pp. 48-57.).

términos y el planteamiento expreso de la necesidad en un sistema de información.⁴⁷ Al principio, la necesidad se mantiene en un plano individual e interno, pero según se hace más latente va tomando forma hasta llegar a un estado denominado “efecto etiqueta” que es cuando el usuario tiene que tiene que expresarla en palabras para poder continuar con el proceso que lo lleve a resolverla. Es justo en esta etapa que la necesidad se materializa en palabras que la representan. Posteriormente, cuando el usuario está ante el sistema de información la necesidad, en su estado final, se traduce en una instrucción de consulta o petición para el Sistema de Recuperación de Información (SRI).

Luego de revisar las distintas definiciones de búsqueda de información, resulta indispensable mencionar la que se usa en este trabajo. Para esta tesis es: un proceso sistemático cuya finalidad es obtener información para resolver una necesidad. Siguiendo a Taylor,⁴⁸ la búsqueda de información inicia con la necesidad no tan consciente, posteriormente, se transforma en una necesidad consciente y luego en una petición, a partir de la cual el sistema de recuperación buscará en las entradas de los registros bibliográficos.

Ahora, tomando como base la definición de búsqueda de información, la búsqueda temática, es una subclase de ésta que tiene los siguientes elementos: un problema o necesidad informativa, un buscador de información, un sistema de búsqueda y un conjunto de resultados. Aún cuando tiene los mismos elementos presentes en otras opciones de exploración (autor, título, clasificación) esto no implica que podamos equipararla con los otros tipos de búsqueda, ya que cada una conlleva una necesidad específica (localizar un autor, una obra de la que se tiene el título, libros agrupados en cierta clasificación, etc.).

En la búsqueda por materias la necesidad de información tiene relación con un tópico en el que el usuario puede estar interesado por distintos motivos, como ser un tema que se tratará en clase, querer saber más al respecto, ser un tema polémico, etc. A diferencia de una exploración por título donde el usuario tiene muy claro que necesita información contenida en una obra específica en la búsqueda por tema éste se mueve a través de las áreas del conocimiento.

⁴⁷ Para la construcción de las etapas me baso en Robert Taylor (Taylor, Robert. “Question negotiation and information seeking in libraries” en *College and research libraries* (Chicago: American Library Association, vol. 29, núm.3, 1968), pp. 178-194.), quien habla de una necesidad visceral, interna; una necesidad consciente dentro del cerebro; una necesidad formalizada, en términos, y una necesidad comprometida donde hay una demanda de información.

⁴⁸ *Idem.*

Cuando se quiere obtener información por temas lo ideal es que la necesidad se encuentre en un estado consciente en el que el usuario sabrá, por ejemplo, que quiere una obra que trate sobre los descubrimientos de especies de ranas.⁴⁹ Luego, el gran paso es que la necesidad sea desplazada del estado consciente al formalizado. Así, el paso de la simple descripción de la necesidad a su expresión en términos comprensibles para el sistema es el asunto central en las búsquedas temáticas. Al respecto Fernández Molina dice que:

La formulación es el punto crítico del proceso. A partir de la información recopilada se identifican y se seleccionan los conceptos e ideas que permiten construir un foco [...] esto es, una perspectiva centrada del tema, que se convierte de alguna forma en algo más personalizado.⁵⁰

Una vez que se tiene la formulación de la consulta es necesario que el usuario haga una petición formal al sistema. La búsqueda por temas implica que el usuario se coloque frente al sistema de recuperación de información, sea capaz de acceder al campo temas y a partir de las palabras seleccionadas cuestione al sistema. J. López Yepes, quien tiene una definición de la búsqueda temática centrada en la forma como operan los sistemas de recuperación dice que el proceso implica:

Una recuperación de referencias bibliográficas o de documentos a partir de los encabezamientos de materia que les han sido asignados. Por lo general se procede interrogando al sistema a partir de una o varias palabras clave que forman la materia, preferentemente más de una cuando son muy comunes -como ocurre con la palabra información- para evitar el ruido. Normalmente, el sistema ofrece como resultado un listado de materias que corresponden a la búsqueda ordenado alfabéticamente por el encabezamiento principal de materia. A partir de esa lista será posible seleccionar la materia objeto de búsqueda.⁵¹

La búsqueda temática hasta mediados de los ochentas era de las opciones más usadas en los catálogos de bibliotecas, aunque distintos estudios la caracterizaban como complicada⁵² debido al desuso de los índices como herramientas de búsqueda

⁴⁹ Si el estado de la necesidad no es de “necesidad consciente” el usuario deberá trabajar más hasta lograrlo, pues no es posible el paso de una necesidad inconsciente a una formalizada.

⁵⁰ Fernández Molina, Juan Carlos y Félix de Moya Anegón. *Los catálogos de acceso público en línea: el futuro de la recuperación de información bibliográfica* (Granada: Asociación andaluza de bibliotecarios, 1998), p. 77.

⁵¹ José López Yepes, ed. *Diccionario enciclopédico de... Op.cit. s.v.* Búsqueda temática

⁵² Los catálogos de bibliotecas están diseñados para acceder a la información por temas de distintos modos, el más sencillo es a través de los índices temáticos en los que el usuario pueden introducir una palabra clave que le permita entrar al listado e ir navegando (browsing) hasta localizar el tema que más se ajuste a su necesidad. No obstante, muchas bibliotecas no ponen a disposición de sus usuarios los índices o bien los usuarios no saben sobre esta opción de búsqueda y optan por la búsqueda mediante interrogación (query). Estudios que analizan las diferencias entre la búsqueda por browsing y query, han concluido que la búsqueda por navegación es más sencilla para los usuarios y, además, los hace entrar en contacto con el lenguaje del sistema, mientras que la segunda es más complicada, empero, la más usada.

(Browsing), el uso generalizado de la búsqueda por interrogación (Query) y el poco conocimiento que los usuarios tiene sobre el lenguaje empleado en la indización.

El desconocimiento del lenguaje de indización por parte del usuario, directamente no tiene que ver con la catalogación, sino más bien con la instrucción. Al menos en México, en muchas bibliotecas hay una separación entre el área de catalogación y el área de servicios. Los bibliotecarios de procesos técnicos generalmente no son quienes instruyen a los usuarios en el uso del catálogo. Esta tarea recae en el bibliotecario de servicios que constantemente lleva a cabo actividades de formación de usuarios, empero, muchas veces en sus planes de formación no incorpora instrucción sobre el uso de lenguajes documentales para auxiliar a los usuarios en las búsquedas por materias. Este hecho impacta de manera directa en la recuperación, ya que el usuario frecuentemente desconoce que en la organización por temas se usa un lenguaje controlado y desconoce los principios de éste. La instrucción en este aspecto parece necesaria para sacar mayor provecho de la catalogación por temas.

1.2.2. La búsqueda por temas en catálogos de bibliotecas

A mediados de la década de los setentas muchas bibliotecas empiezan a sustituir sus catálogos en tarjetas por catálogos automatizados. De esos años a la fecha, los catálogos han sufrido transformaciones debidas al avance tecnológico, las investigaciones dirigidas a su mejora y las necesidades de los usuarios. Charles Hildreth⁵³ estableció tres generaciones de catálogos. La primera generación se caracterizó por ser muy parecida a los catálogos en tarjetas, pues sólo ofrecían tres puntos de acceso: autor, título y materia. Fueron catálogos poco flexibles en los que era necesario introducir con exactitud las palabras a buscar para lograr coincidencia con los índices.

La segunda generación de catálogos fue más flexible, porque permitieron la incorporación de elementos de exploración distintos a los tradicionales, como la indagación por palabras clave, brindaron la posibilidad de restringir la búsqueda mediante operadores booleanos, fecha, lengua y lugar de publicación y permitieron realizar búsquedas mediante el cruce de índices. Otro beneficio de esta generación fue que facilitó la ampliación del rastreo a otros elementos del registro, como tablas de contenidos, sumarios, notas de contenido y el texto completo.

⁵³ Charles R. Hildreth. *The online catalogue: developments and directions* (London: The Library Association, 1989), p. 17.

Parece ser que estos catálogos de segunda generación son los que tenemos actualmente en la mayoría de las bibliotecas. En palabras de Luis Villén Rueda:

Sólo unos pocos sistemas han llegado más allá de los catálogos en línea de segunda generación, los cuales han dado lugar a catálogos en línea de tercera generación, con características mejoradas o más sofisticadas. Gracias al desarrollo de la tecnología se incorporan nuevas posibilidades de recuperación, haciéndolos más adaptables a las necesidades de los usuarios de las bibliotecas.⁵⁴

Una desventaja de esta generación de catálogos que tiene impacto en la recuperación temática es que para la extracción de información se basan en la concordancia exacta (exact-matching) entre el término de búsqueda y un dato del registro bibliográfico. Ante este inconveniente algunas bibliotecas, con la finalidad de hacer más efectiva la recuperación, han optado por establecer algoritmos de concordancia parcial (partial-matching) mediante la reducción de los términos a la raíz y la corrección de errores ortográficos, empero, este proceso ha ocasionado poca precisión en los resultados, pues una raíz temática es origen de una gran cantidad de vocablos.

Pocas bibliotecas cuentan con catálogos de tercera generación. Éstos se caracterizan por ser OPACS que tras su diseño o implementación son respaldados por estudios de uso⁵⁵ y están más vinculados a las necesidades de los usuarios. Suelen tener ventajas como la opción de búsqueda en texto libre, sistemas que enlazan (mediante tesauros) las peticiones del usuario con las herramientas de indización de la biblioteca, incorporan bases de datos enriquecidas que guían los términos del usuario hacia las entradas normalizadas, vinculan con la terminología de otros catálogos de bibliotecas y tiene interfaces diseñadas bajo principios de usabilidad.

Los catálogos de tercera generación han superado muchos de los problemas que presentaban sus antecesores. Sin embargo, como señala Villén Rueda, en la mayoría de las bibliotecas académicas mexicanas los catálogos son de segunda generación y hay instituciones que aún cuentan con catálogos de primera generación.

⁵⁴ Luis Villén Rueda. *La indización y...* *Op.cit.* p. 83. En lo que respecta al catálogo de la Biblioteca Daniel Cosío Villegas, éste se encuentra en los de segunda generación pues: ha superado a los de primera generación que fueron muy similares a los catálogos en tarjetas; reúne las características de la segunda generación (búsqueda por palabras clave, búsquedas mediante operadores booleanos, posibilidades de truncamientos, opciones para el cruce de índices en la búsqueda avanzada y desde su interfaz permite restringir la búsqueda por fecha, lengua, año y formato de materiales), y no puede incluirse en los de tercera generación que tienen correctores ortográficos, interfaces que pueden ser altamente personalizadas (opciones a cambios de colores, plantillas, etc.) y ligas a bases multi-tesauros.

⁵⁵ Para mayores referencias en torno a estudios sobre catálogos de bibliotecas y las mejoras que se hicieron en éstos luego de las investigaciones véase a Lancaster (Lancaster, F. W. et. al. "Identifying barriers to effective subject access in library catalogs" en *Library Resources and Technical Services* (Chicago: American Library Association, vol 35, núm. 4, 1991), pp. 377-391.).

Los catálogos de primera y segunda generación han sido definidos como difíciles de utilizar, sobre este aspecto, Christine Borgman⁵⁶ señala que son complicados de usar, porque si bien han sido estudiados como sistemas de recuperación, no se han detectado sus problemas en el uso real y su diseño no se ha ligado a la conducta de búsqueda de los usuarios. Esto se debe a que son catálogos diseñados por empresas comerciales y profesionales ajenos a las ciencias documentales y a la bibliotecología que desconocen cómo se lleva a cabo la organización de la información, las necesidades de los usuarios y la vinculación entre unos y otros.

En lo que respecta a la búsqueda por materias, los catálogos de bibliotecas de segunda generación ofrecen posibilidades bastante limitadas. A eso se debe que buena parte de los estudios orientados a resolver los problemas del catálogo sean más bien estudios para solucionar problemas del rastreo por temas. En este tono, Juan Carlos Fernández Molina sobre la búsqueda por temas en los OPACS dice que:

Es al área más necesitada de mejora y donde por tanto se han centrado la mayoría de las investigaciones al respecto [...] La razón básica se encuentra en que, como numerosos estudios han demostrado, la mayoría de las búsquedas de los usuarios de los OPAC son de naturaleza temática y, además, son las que presentan mayores problemas.⁵⁷

En las bibliotecas mexicanas resulta necesario llevar a cabo investigaciones sobre la recuperación por materias, pues es uno de los puntos de acceso más importantes. Otro motivo por el que es necesario estudiarla es su complejidad, esto es, su supuesta dificultad para ser llevada a cabo. Sobre el asunto, los estudios señalan que la búsqueda por temas presenta los siguientes inconvenientes:

- a) Problemas para localizar el campo de búsqueda
- b) Dificultad en la elección de términos para realizar la consulta
- c) Búsquedas que no recuperan documentos
- d) Búsquedas que recuperan demasiados documentos
- d) Los resultados obtenidos no corresponden con lo que el usuario esperaba.

Exceptuado el problema uno que concierne a la interfaz del catálogo, los demás problemas se deben a lo que Ray Larson caracteriza como poco conocimiento del usuario sobre la indización. El autor indica que:

El desconocimiento de los sistemas mediante los cuales las bibliotecas asignan encabezamientos de materias son causa de búsquedas cuyos resultados son ninguno,

⁵⁶ Borgman, Christine. "Why are online catalogs still hard to use" en *Journal of the American Society for Information Science* (Nueva York: John Wiley & Sons vol. 47, núm. 7, 1996), pp. 493-503.

⁵⁷ Juan Carlos Fernández Molina y Félix de Moya Anegón. *Los catálogos de... Op.cit.* pp. 45-46.

muy pocos registros o búsquedas que no se ajustan a lo que el usuario tenía en mente. El mismo desconocimiento de las herramientas controladas de indización origina dificultades conceptuales en la formulación de las búsquedas. Otro aspecto es que los usuarios en la formulación de búsquedas introducen términos muy amplios que en la recuperación ofrecen demasiados registros.⁵⁸

En las bibliotecas, los usuarios de los catálogos en línea son un grupo heterogéneo con características y necesidades de información muy distintas. Mientras unos son personas muy experimentadas en la búsqueda de información, otros son prácticamente novatos. Los primeros, debido a su experiencia, obtienen resultados más satisfactorios, mientras que los segundos se enfrentan a dificultades. Por tal motivo, al segundo grupo el catálogo debe proveerle ayudas. En este tono, J. A. Adams dice que: “El OPAC ha de jugar un papel democratizador en el acceso a la información de todo tipo y en todos sus formatos y ha de adaptarse a los nuevos entornos de la información para satisfacer las demandas de los usuarios.”⁵⁹ También Villen Rueda quien está a favor de diseñar un catálogo más dirigido al usuario dice que “es momento de dejar atrás los tiempos en que el catálogo era producto de la fascinación bibliotecaria hacia la tecnología y resultaba más atractivo para las bibliotecas y los bibliotecarios que para los usuarios.”⁶⁰

La búsqueda por materias puede ser de mucha utilidad cuando se está buscando información sobre un tópico específico, es un punto de acceso fundamental a la obtención de información y es una actividad en la que se invierten recursos. De ahí la necesidad de indagar sobre su estado en las bibliotecas, elaborar un diagnóstico y de ser éste negativo implementar acciones para sacarle el mayor provecho y ponerla al servicio de los usuarios.

1.3. El lenguaje en la recuperación por materias

1.3.1. El lenguaje de la indización

El lenguaje de la indización puede ser libre o controlado. Sin embargo, en la organización de información en bibliotecas se da prioridad al controlado como herramienta de indización y al referirse al lenguaje de indización la mayoría de las obras de referencia de la disciplina asocian lenguaje de indización con lenguaje controlado.

⁵⁸ Ray Larson. “The decline of subject searching: long-term trends and patterns of index use in an online catalog” en *Journal of the American Society for Information Science* (Nueva York: John Wiley & Sons, vol. 42, núm. 3, 1991), pp. 197-215.

⁵⁹ J.A. Adams. “The computer catalog: a democratic or authoritarian technology” en *Library Journal* (Nueva York: R. R. Bowker, vol. 113, núm. 2, 1988), pp. 31-36.

⁶⁰ Luis Villen Rueda. *El acceso temático...Op.cit.* p. 91.

Las definiciones sobre el lenguaje de indización son muy parecidas, lo que lleva a pensar que existe consenso en torno a su concepto y su función. J. M Reitz dice que es:

Un lenguaje artificial consistente de encabezamientos temáticos o descriptores de contenido seleccionados para facilitar la recuperación de información sirviéndose de puntos de acceso en un catálogo o índice, incluye entradas autorizadas y reglas que gobiernan la forma de las entradas, sintaxis, etc.⁶¹

Una segunda definición nos dice que es:

Un lenguaje documental que permite representar el contenido de los documentos para su recuperación [...] Posibilita la entrada de documentos al sistema bibliográfico atendiendo a su contenido. Igualmente desempeña una labor fundamental en el momento de la recuperación.⁶²

El lenguaje artificial, desde la perspectiva de Pierre Lerat,⁶³ es aquel en el que el vocabulario, sintaxis, semántica y pragmática están limitadas y definidas. A diferencia del lenguaje natural, es manipulado para finalidades específicas y es construido con una finalidad pragmática que establece la disciplina o el dominio al que sirve. Los términos que incluye son un subconjunto de la lengua natural y para su elección pasan por un principio de garantía literaria o garantía de uso. Esto implica que son tomados de fuentes literarias y, además, se verifica su uso en la comunidad.

El empleo de vocabularios controlados en la indización responde a dos finalidades:

- a) Asegurar la representación consistente de las materias, tanto en indización como en la recuperación. Esto se consigue mediante el control de la sinonimia y la homografía. Será más fácil que los indizadores entre sí y entre éstos y los que hacen la búsqueda, coincidan en los términos escogidos para representar un concepto si son seleccionados de una lista previa que si utilizan el lenguaje natural. El control de la sinonimia suele hacerse con el envío de use o ver de los términos no utilizados al sinónimo utilizado (pájaro use aves) y el control de la homografía se establece añadiendo el término un calificador entre paréntesis. Por ejemplo, diafragma (anticonceptivo).
- b) Facilitar la realización de búsquedas amplias sobre una materia, relacionando los términos mediante relaciones de jerarquía y asociación. En la recuperación se podrán identificar fácilmente todos los términos necesarios para ampliar la búsqueda sobre una materia. Si el usuario encuentra en el vocabulario que el término que le interesa, rapaces, tiene un genérico (aves) o un relacionado (ornitología) le sugiere la posibilidad de no buscar sólo sobre el primer término y ampliar la búsqueda con los términos relacionados.⁶⁴

⁶¹ Joan M. Reitz. *Dictionary for Library... Op.cit. s.v. Lenguaje de indización.*

⁶² José Lopez Yepes, ed. *Diccionario enciclopédico de... Op.cit. s.v. Lenguaje de indización.*

⁶³ Lerat, Pierre. *Las lenguas especializadas* (Barcelona: Ariel, 1997), p. 85.

⁶⁴ José Lopez Yepes, ed. *Diccionario enciclopédico de... Op.cit. s.v. Lenguaje controlado (Objetivos).*

En la indización se da prioridad al lenguaje controlado sobre el lenguaje libre con la finalidad de asegurar la consistencia en la asignación temática y facilitar la recuperación. No obstante, el lenguaje natural tiene elementos como la sinonimia y la polisemia que el lenguaje controlado debe contemplar, es por ello que los sinónimos son incluidos como referencias, con instrucciones de *véase* o *útese* el asiento autorizado. Por ejemplo, si un usuario busca el término “Pájaros” y la biblioteca en la indización prefiere el término “Aves”, debe haber una referencia que al usuario le permita saber que el término aceptado es “Aves”. De tal forma, se asegura que quienes hayan buscado “Pájaros” sean remitidos al encabezamiento admitido. Gran parte del trabajo de las bibliotecas está dirigido a establecer relaciones entre los distintos sinónimos de una palabra y los términos autorizados, mediante la creación y constante enriquecimiento del catálogo de autoridades temáticas. Empero, sigue siendo tema de discusión la posibilidad de que el lenguaje controlado pueda incorporar en registros de autoridad todas las manifestaciones de un concepto.

Los lenguajes de indización más usados en bibliotecas son las listas de encabezamientos y los tesauros. Las listas de encabezamientos son lenguajes pre-coordinados y reciben este nombre porque proporcionan al indizador ya vinculados los símbolos que constituyen un tema. Una de las listas más conocidas es LCSH,⁶⁵ elaborada por Library of Congress, mientras que en el contexto hispanoamericano hay otras: la lista de *Encabezamientos de materia LC en español*,⁶⁶ la *Lista de Encabezamientos de Materia para Bibliotecas (LEMB)*,⁶⁷ la *Lista de encabezamientos de materia de la red de bibliotecas del CSIC*⁶⁸ y la *Lista de encabezamientos de materia*⁶⁹ de Gloria Escamilla.

Según José María Izquierdo Arroyo,⁷⁰ las LEM aparecieron en una época en la que predominaba la preocupación por clasificar los documentos física y temáticamente

⁶⁵ United States. *Library of Congress Subject Headings* (31 ed. Washington: Cataloging Distribution Service, Library of Congress, 2009.).

⁶⁶ United States. *Encabezamientos de materia LC en español* (Washington: Library of Congress, 2010. (DE, 25 de noviembre, 2010: <http://lcsch-es.org>)).

⁶⁷ Biblioteca Luis Angel Arango. *Lista de encabezamientos de materia para bibliotecas* (Santafé de Bogotá: R. Eberhard, 1998.).

⁶⁸ España. Consejo Superior de Investigaciones Científicas. Unidad de Coordinación de Bibliotecas. *Lista de encabezamientos de materia de la red de bibliotecas del CSIC* (Madrid: Consejo Superior de Investigaciones Científicas, 2000.).

⁶⁹ Escamilla González, Gloria. *Lista de encabezamientos de materia* (México: UNAM, Instituto de Investigaciones Bibliográficas, 1978.).

⁷⁰ José María Izquierdo Arroyo y Luis Miguel Moreno Fernández. “Listas de encabezamientos de materia y thesauri en perspectiva comparada” en *Documentación en ciencias de la información* (Madrid: Universidad Complutense, núm. 17, 1994), pp.287-310.

para ser localizados. Pero la explosión de la información, la diversificación temática de los documentos y las exigencias actuales en función de recuperación desbordaron la capacidad de aquéllas para representar y recuperar materiales. Fue entonces cuando se hicieron más necesarias otras herramientas de indización y se elaboraron tesauros.

Éstos son listados de término sobre disciplinas específicas. Son lenguajes post-coordinados, lo que les permite más posibilidades en la combinación de estructuras, suelen centrarse en un campo concreto del conocimiento y se nutren de un lenguaje más especializado y técnico que las listas de encabezamientos. En la recuperación suelen ser más efectivos, en virtud de que tienen una mayor riqueza léxica y posibilidades de combinación entre términos.

Durante los años de convivencia de los lenguajes precoordinados y poscoordinados ha habido enfrentamientos entre los defensores de unos y otros respecto a cuál herramienta es mejor. Pero, no se ha logrado llegar a conclusiones al respecto, empero, es importante decir que la herramienta que una biblioteca emplee en la indización será determinante cuando el usuario, pues mientras una lista de encabezamientos es más general, en el sentido de que contiene términos de distintas disciplinas, esta característica implica que recoge menor cantidad de vocablos de cada una de ellas en comparación los tesauros que están dirigidos a un área específica del conocimiento y cuentan con más amplitud terminológica del campo de conocimiento que tratan.

1.3.2. El lenguaje en la búsqueda temática

El lenguaje de la búsqueda es el del usuario, es decir, un subconjunto del lenguaje natural con el que la gente habla y escribe. López Yepes lo define como “un lenguaje humano cuyas reglas derivan del uso.”⁷¹ En éste no hay esfuerzo extra, aparte de lo que implica el dominio de la lengua, para limitar o definir el vocabulario, la sintaxis o la semántica. En documentación el lenguaje natural es usado de tres modos: en las palabras de los títulos, encabezados y partes del documento; el uso de frecuencias de términos extraídos de cualquier parte del documento para la representación (indización automática) y en las palabras empleadas por los usuarios para formular una consulta en un sistema de recuperación de información.

El lenguaje natural y sus subconjuntos se componen de dos tipos de palabras: con significado y sin significado. Las palabras con significado son aquellas que tienen

⁷¹ José Lopez Yepes, ed. *Diccionario enciclopédico de...Op.cit. s.v.* Lenguaje de indización.

sustancia, es decir, que refieren a conceptos y tienen un referente, mientras que las carentes de significado suelen ser los artículos, preposiciones y conjunciones.

Una cualidad del lenguaje natural es que no realiza esfuerzos por mantener el control semántico. Rao Aluri dice que en este lenguaje “un concepto puede ser representado por dos o más términos sin que existan ligas entre éstos que sean rescatadas en la recuperación de información.⁷² Esta característica lo opone al lenguaje controlado, pues éste busca controlar la sinonimia y la polisemia con la finalidad de eliminar la ambigüedad en la recuperación de documentos.

El vocabulario no controlado puede contener sinónimos e incluso formas variantes en la escritura. Si es empleado en la indización, su uso es ventajoso ya que el esfuerzo para indizar un documento tiende a ser menor que en la indización bajo términos controlados. Mientras que en el lenguaje natural se pueden tomar las palabras de título o del contenido del documento como términos de indización, en los lenguajes controlados es necesario adaptarse a un listado autorizado.

Durante muchos años el lenguaje natural fue usado en la representación, luego, vino un periodo de abandono debido al surgimiento de los lenguajes controlados, pero, en los últimos años, gracias a las posibilidades actuales de almacenamiento de grandes cantidades de texto en el ordenador, la mejora de las ayudas de búsqueda en las bases de datos y el desarrollo de procedimientos de indización automatizada, ha resurgido como herramienta de indización, ya sea solo o en combinación con algún lenguaje controlado.

Si bien emplear el lenguaje natural resulta riesgoso en la catalogación temática por su dispersión terminológica, se debe tener en cuenta que es el lenguaje con el cual buscan los usuarios. Es por ello que si una biblioteca indiza con lenguajes controlados debe buscar que las peticiones en lenguaje natural recibidas en el catálogo de una u otra manera encuentren respuesta en el lenguaje de la indización.

1.3.3. El lenguaje de indización frente al lenguaje de búsqueda de información

El lenguaje es, por excelencia, la principal herramienta de comunicación y en los sistemas de recuperación se torna un vínculo entre las necesidades del usuario y las representaciones de los materiales. Según Rao Aluri:

⁷² Rao Aluri; Alasdair D. Kemp y John J. Boll. *Subject analysis in online catalogs* (Englewood, Colorado: Libraries Unlimited, 1991), p. 38.

El lenguaje tiene influencia sobre el proceso de recuperación de información de tres modos: primero, las necesidades de información de los usuarios son expresadas como preguntas que toman forma lingüística; segundo, la información en los documentos es, en gran medida, expresada en palabras; tercero, el lenguaje, en sí mismo, adquiere una función importante al momento de recuperar información”.⁷³

Los sistemas de recuperación de información presentan dos momentos, que son las dos caras de la organización de la información: por un lado, está el momento de entrada (input) que corresponde a la indización y, por otro lado, hay un momento de salida (output) que atañe a la recuperación. Esta función de los sistemas de recuperación quedaría como sigue:

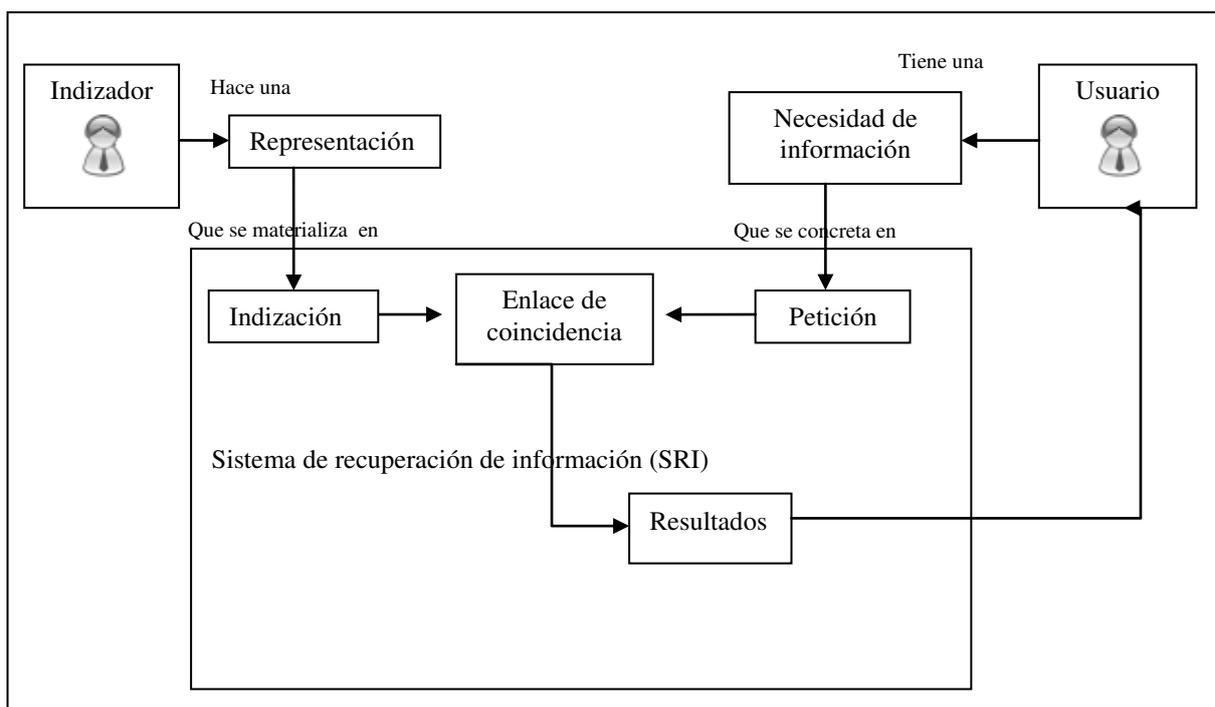


Fig.4. Función del sistema de recuperación en la búsqueda por temas

Como se aprecia en el esquema anterior, el sistema, en este caso catálogo de la biblioteca, abarca desde la representación de los documentos hasta la recuperación de éstos. Pareciera que la representación, concretada en la indización, va en un sentido mientras que la recuperación corre en el sentido inverso; no obstante, ambas actividades son un todo. Hasta principios de la década de los setentas, la representación y la recuperación se había visto como actividades desligadas, pero luego de la automatización bibliotecaria y el surgimiento de moderno sistemas de recuperación se han visto como dependientes. Incluso, recientemente, a partir paradigma de indización

⁷³ *Ibidem.* p. 28.

centrada en el usuario, se empieza a postular que deben ir de la mano, vistas desde un enfoque sistémico,⁷⁴ cuyo fin es responder a las necesidades del usuario.

La búsqueda temática tiene diferentes etapas. Las que se relacionan con el lenguaje en los sistemas de recuperación son: la formulación de interrogación, momento en el cual el usuario traduce su necesidad en términos y la petición de búsqueda, momento en que el usuario teclea las palabras en el sistema y envía la petición. Viene entonces una fase determinante: la función de enlace o coincidencia que el motor de rastreo tienen que establecer entre la petición y la indización. Esta función desde el punto de vista del sistema de recuperación parece muy sencilla, ya que se resume al hecho de que un motor de búsqueda detecte coincidencias entre la petición e indización o bien que a partir del término solicitado establezca ligas a otros términos. Empero, esta coincidencia que pareciera sencilla no siempre es posible, pues se está intentando hacer coincidir el lenguaje del usuario- lenguaje natural- con un lenguaje de indización- lenguaje controlado. Éstos, según la teoría de recuperación de información, deben tener empalme en tanto que uno se deriva del otro, pero dadas sus características, objetivos y distanciamiento algunas veces no coinciden.

La discusión sobre las diferencias entre el lenguaje de indización y el lenguaje del usuario han sido ampliamente tratadas en lo que respecta a cómo estos conviven en los modernos SRI. Heting Chu⁷⁵ dice que en esta discusión se pueden distinguir cuatro etapas:

La primera, refiere al momento anterior a la aparición de los lenguajes controlados. Entonces, el lenguaje natural era el único usado para representar y recuperar información. Fue en ese tiempo cuando se observó que su uso implicaba problemas, ya que en su uso intervenían casos de homonimia, polisemia y sinonimia que lo hacían ambiguo. Un segundo momento, marcado por la introducción del vocabulario controlado, se caracterizó por el surgimiento de herramientas tales como

⁷⁴ El enfoque sistémico es una teoría económica que trata de comprender el funcionamiento de la sociedad desde una perspectiva holística e integradora, en donde lo importante son las relaciones entre los componentes. En años recientes en enfoque sistémico se ha extendido a otras disciplinas, pues destaca las relaciones que existen entre los componentes de un sistema para el logro de un fin común. Si se aplica este enfoque a los sistemas de recuperación de información se observa que éste tiene partes que lo constituyen: bases de datos, alimentadores, motores de búsqueda, interfaz, peticiones de usuario, etc. Cada uno de sus elementos tiene una función dentro del sistema dirigida al logro de un objetivo común: posibilitar la recuperación de información. Tomando como base este objetivo, podría definírseles mejor e integrárseles para el logro de la meta final.

⁷⁵ Chu, Heting. *Information, representation and retrieval in the digital age* (Meford, NJ. : Information today, 2003), p. 52.

esquemas de clasificación, tesauros y listas de encabezamientos. Fue cuando surgió con mayor énfasis la idea de “lenguaje controlado vs lenguaje natural” y se afirmaba que alguno de ellos, de preferencia el controlado, debía suprimir al otro.

La tercera etapa se caracterizaría por el resurgimiento del lenguaje natural, que fue posible a raíz de que los catálogos permitieron la búsqueda por palabra clave y surgieron sistemas de recuperación más efectivos capaces de buscar en todo el documento. El resurgimiento del lenguaje natural no implica la desaparición del lenguaje controlado, aún cuando muchos especialistas en organización de la información han cuestionado la utilidad del lenguaje controlado y han especulado sobre su desaparición frente a nuevas formas de indización (social y automática).

La cuarta etapa de los lenguajes de representación se caracteriza por la implementación de interfaces de lenguaje natural en las que la herramienta principal sigue siendo el lenguaje controlado, pero es invisible para el usuario, es decir, el usuario realiza una búsqueda en lenguaje natural y el sistema tiene la habilidad de vincular su petición con los encabezamientos aceptados.

Sobre en qué etapa nos encontramos, Heting Chu indica que:

Aunque los límites entre las diferentes áreas no son demasiado obvios y no hay un corte claro entre una y otra, la representación y la recuperación hoy en día, aparentemente, ha ya pasado las primeras dos eras y se expande hacia la tercera y cuarta en términos de lenguajes de representación y recuperación de información.⁷⁶

La coexistencia del lenguaje natural y el lenguaje controlado en un sistema de recuperación se da de distintos modos, pues ambos pueden ser utilizados tanto en la representación como en la recuperación. En aquellos sistemas que indizan en lenguajes controlados se prefiere el uso de éste, porque años atrás se detectaron problemas en el empleo del lenguaje natural que se intentaron superar mediante la aplicación de vocabularios controlados. Los inconvenientes del lenguaje natural que se intentaron superar con el empleo del lenguaje controlado fueron: la sinonimia, polisemia y homonimia.

Tanto el lenguaje natural como el lenguaje de indización tienen una finalidad comunicativa. Sin embargo, el lenguaje de indización tiene una segunda función: servir para la organización de conocimiento. En virtud de esta meta, es un subconjunto del lenguaje natural dotado de características semánticas, sintácticas y léxicas particulares que lo distancian del conjunto del que se desprende.

⁷⁶ *Idem.*

Una diferencia fundamental entre ambos lenguajes es la sintaxis: el lenguaje natural tiene una sintaxis rica que corresponde a una gramática social, mientras que el lenguaje controlado tiene una sintaxis particular orientada hacia el enlace de facetas que enlazadas resultan útiles para la descripción de los materiales. Esta sintaxis resulta lógica y comprensible para el representador, pues su formación profesional incluye los aprendizajes teóricos de la sintaxis de los lenguajes controlados; no obstante, es casi desconocida para los usuarios quienes en sus búsquedas por temas emplean la sintaxis propia del lenguaje natural.

La cobertura léxica es otro aspecto a comparar entre el lenguaje natural y el controlado. Mientras que el primero es amplio y presenta gran cantidad de vocablos para expresar un tema; el segundo no engloba todas las palabras de una lengua, sino más bien selecciona términos que puedan ser objeto de topicalización. En los lenguajes predomina la estructura nominal, estrechamente vinculada con la expresión de referentes. En el mismo plano léxico, la actualización es otro elemento diferenciador mientras que el lenguaje natural se actualiza por sí mismo; el controlado necesita ser actualizado para incorporar nuevos términos. La incorporación de neologismos en los lenguajes controlados implica cierto tiempo durante el cual se observa el comportamiento del término, su consolidación en la lengua y se valora si puede adherirse como encabezamiento de materia, mientras esto sucede, la producción literaria sobre el tema queda sin espacio de clasificación.

El hecho de que confluyan ambos lenguajes en un SRI implica la priorización de uno sobre el otro y, en la mayoría de los casos, es el usuario quien tiene que adaptarse al lenguaje del sistema. Así, al buscar por temas, debe formular una consulta comprensible para éste. Un catálogo en línea no podría responder a consultas como: ¿Tienes un libro sobre la descripción física de Fresno California? Si el usuario mediante el uso del catálogo quiere obtener materiales que respondan a necesidades de información planteadas debe transformar su necesidad en una consulta comprensible para la máquina. Su necesidad debe desplazarse del plano conceptual al plano del símbolo y aún más, símbolos comprensibles para el sistema. Bajo los principios semánticos, léxicos y sintácticos del lenguaje del catálogo, su necesidad debe formularse como: Fresno, California- Descripción y viajes.

La transformación de la simple descripción de la necesidad en una formulación de búsqueda, según J. C. Fernández Molina implicaría que:

Para conseguir la información adecuada que satisfaga sus necesidades de información, los usuarios deben formular su búsqueda mediante términos que sean entendibles por el catálogo, es decir, tiene que hacer coincidir su lenguaje con el que utiliza el sistema en las descripciones bibliográficas. Por tanto, hay un lenguaje estándar conocido y aceptado, el del sistema, y para ser un usuario competente del sistema es necesario conocer dicho lenguaje.⁷⁷

En gran medida, para que una recuperación sea exitosa, dice Rao Aluri “debe haber coincidencia entre la terminología de indización empleada en el sistema y la consulta del usuario.”⁷⁸ Sin embargo, uno de los principales problemas que se presenta en la recuperación de información se relaciona con esta falta de coincidencia entre el lenguaje natural empleado por el usuario en sus búsquedas y el lenguaje controlado usado por el indizador para representar los documentos.

Sobre el problema varios autores han investigado. Al respecto, Karen Antell y Jie Huang dicen que la exploración por temas en los catálogos en línea tiene un carácter complejo por “la dificultad que implica para el usuario seleccionar los términos adecuados que debe introducir para localizar documentos.”⁷⁹ También sobre el asunto, Mari Carmen Marcos señala que:

La recuperación de información relevante en los OPAC resulta difícil para los usuarios cuando se trata de búsquedas por materias, ya que en la mayoría de los casos los usuarios desconocen los descriptores o términos empleados por los catalogadores para indizar los documentos.⁸⁰

La probabilidad de coincidencia entre el lenguaje de indización y el lenguaje del usuario en una consulta temática, según investigaciones llevadas a cabo parece ser muy baja. En el mejor de los casos el porcentaje de coincidencia oscila entre el 20 y 25%.⁸¹ El hecho implica un problema, pues el enlace entre la petición y las entradas de la base de datos es fundamental ya que de él depende la recuperación de materiales. Al no haber esta función de enlace se obtienen búsquedas cuya recuperación es cero.

Los sistemas de recuperación exigen al usuario una buena formulación de su demanda de información. Esto, según Belkin, es pedir demasiado, ya que “para expresar sus necesidades de información tienen que describir algo que no saben con

⁷⁷ Juan Carlos Fernandez Molina y Felix de Moya Anegón. *Los catálogos de... Op.cit.* p. 49.

⁷⁸ *Ibidem.* p.54.

⁷⁹ Karen Antell y Jie Huang. “Subject searching success...*Op.cit.*

⁸⁰ Marcos, Mari Carmen. *Mejoras en la consulta...Op.cit.*

⁸¹ Véase Marcia Bates (Bates, M. J. “System meets user: problems in matching subject search terms” en *Information processing and Management* (Oxford: Pergamon, vol. 13, núm. 6, 1977), pp. 367-375.); Marilyn Ann Lester (Lester, Marilyn Ann. *Coincidence of user vocabulary and Library of Congress subject headings: experiments to improve subject access in academic library online catalogs* (Urbana, Ill.: University of Illinois at Urbana Champaign, 1988.); Karen Markey (Markey, Karen. *Online catalog use: results of surveys and focus group interviews in several libraries* (Washintong: OCLC, 2000 (DE, 13 de Agosto, 2010: <http://library.oclc.org>)).

certeza.”⁸²Tienen en su mente una necesidad que pueden describir, pero en su intento por encontrar las palabras que la representen puede ser que su representación no coincida con la de la biblioteca. Así, para muchos usuarios la tarea de buscar documentos por materia resulta una actividad “fatigosa”⁸³ y luego de introducir varias palabras, de modo erróneo concluyen que el catálogo, y por ende la biblioteca, no contienen documentos que respondan a su búsqueda por tema. Esta idea es la mayoría de las veces errónea, pues en un gran porcentaje de búsquedas por temas la falta de recuperación se genera ante la falta de enlace entre las peticiones del usuario y la indización. Ante este problema, las bibliotecas deben emprender estudios exploratorios sobre el estado de la recuperación temática que ayuden para dar un diagnóstico de eficiencia en la recuperación temática y den luces hacia una futura solución del problema.

1.4. Conclusión

Los catálogos de bibliotecas tienen dos momentos estrechamente vinculados: el de entrada (indización) y el de la salida (búsqueda por temas). En la indización un catalogador representa un documento mediante el empleo de un lenguaje específico (en el caso de la BDCV lenguaje controlado). En la salida, un usuario mediante el empleo de su lengua natural busca por temas para resolver una necesidad de información. Son actividades que parecieran desligadas, pero no es así; más bien son complementarias y dependientes.

En la recuperación temática en catálogos de bibliotecas intervienen muchos elementos de la indización y de la búsqueda por temas que repercuten en los resultados obtenidos por los usuarios. Por tanto, toda biblioteca y toda investigación al respecto debe delimitar tanto a nivel conceptual como a nivel de proceso ambos momentos y analizar si la organización de la información está siendo congruente a las opciones de recuperación que se cree se están ofreciendo. Es necesario determinar si la indización está siendo útil y si coincide con la manera como busca el usuario.

⁸² N. J. Belkin, R. N. Oddy y H. M. Brooks. “ASK for information retrieval: part 1. Background and theory en *Journal of documentation* (Londres: Association of Special Libraries and Information Bureaux, vol. 38, núm. 2, 1982), pp.61-71.

⁸³ Las búsquedas por temas en la investigación de Jeffrey Beall (Beall, Jeffrey. “Search Fatigue” en *American libraries* (Chicago, Ill.: American Library Association, vol. 38, núm. 3, 2007, pp. 46-50.) fueron descritas como fatigosas, en el sentido de que el usuario buscaba usando un término y no encontraba información, empleaba otro y tampoco, así sucesivamente hasta terminar cansados sin obtener lo que realmente querían.

En la recuperación por temas el lenguaje es un elemento importante, pues se torna un vínculo de comunicación entre el usuario y las colecciones de la biblioteca. El catálogo lleva a cabo la función de intermediario y en él confluyen el lenguaje controlado usado en la indización y el lenguaje natural de la interrogación. Tanto el lenguaje de la indización como el lenguaje de la búsqueda (del usuario) son subconjuntos del lenguaje natural, lo que podría suponer cierto nivel de correspondencia en tanto que se derivan de un mismo grupo; no obstante, en función de sus particularidades se alejan hasta el punto de que en un sistema de recuperación ya no se comunican, es decir, no tienen posibilidad de enlace lo que impacta en la recuperación por temas.

El estudio teórico de los elementos presentes en la indización y la búsqueda por temas debe continuarse, pero también deben emprenderse más investigaciones al respecto en virtud de que los resultados obtenidos servirán para implementar acciones que mejoren las actividades y las alineen para que sean más útiles a las bibliotecas y los usuarios.

Capítulo 2. Encuadre institucional

2.1. El Colegio de México y la Biblioteca Daniel Cosío Villegas

El Colegio de México (COLMEX) es un organismo descentralizado del gobierno mexicano cuyos objetivos son la investigación y la formación académica en los ámbitos de las ciencias sociales y las humanidades. Fue fundado en el año de 1940 con el nombre La Casa de España, y durante sus primeros años, bajo el gobierno del entonces presidente Lázaro Cárdenas dio asilo a un numeroso grupo de intelectuales procedentes del exilio español. Desde su fundación hasta la actualidad, ya con un segundo nombre: El Colegio de México, la institución se ha destacado en sus labores: “convirtiéndose en uno de los centros de investigación y docencia más destacados en el mundo hispánico”.⁸⁴

Aunque en sus primeros años la prioridad de la institución fue la investigación, la formación académica se vislumbró necesaria. Es por ese motivo, que mediante la dirección de destacables figuras, comenzaron a establecerse los centros de estudio que poco a poco fueron integrándose a la institución hasta conformar un todo. Así:

El Colegio se fue expandiendo hacia nuevas áreas de conocimiento y docencia en posgrado (maestrías y doctorados) [...] el 16 de enero de 1961 se pidió la modificación del Acta Constitutiva de El Colegio para incorporar la fracción f al artículo 1º para agregar a sus objetivos el impartir enseñanzas a nivel universitario, post- profesional o especiales, en las ramas de conocimientos humanísticos y de las ciencias sociales y políticas, creando los órganos apropiados a la realización de estos fines y otorgando los diplomas, títulos y grados correspondientes, de acuerdo con los planes y programas de estudios de la institución que se aprueben.⁸⁵

Actualmente, El Colegio de México integra siete centros de investigación: Centro de Estudios Históricos (CEH); Centro de Estudios Demográficos Urbanos y Ambientales (CEDUA); Centro de Estudios Lingüísticos y Literarios (CELL); Centro de Estudios Económicos (CEE); Centro de Estudios Internacionales (CEI); Centro de Estudios sociológicos (CES) y Centro de Estudios de Asia y África (CEAA). En ellos se llevan a cabo diversas actividades: investigación, vinculación con entidades sociales, colaboración académica con otras universidades, presencia en reuniones académicas, publicación en el área de especialidad (cada centro cuenta con al menos una publicación) y docencia. Sobre el último aspecto, vale la pena destacar que:

El Colegio de México imparte 18 programas docentes de licenciatura, maestría y doctorado; dos son licenciaturas: "Relaciones Internacionales" y "Política y Administración Pública", y el resto son programas de postgrado. Entre ellos se cuentan los doctorados en Historia, en Lingüística, en Literatura Hispánica y en Ciencias Sociales con Especialidad en Sociología, así como las Maestrías en Demografía y en Economía, todos los cuales forman parte del

⁸⁴ El Colegio de México. *El Colegio de México* (México: El Colegio de México: 2010. (DE, 31 de Agosto, 2010: <http://www.colmex.mx>)).

⁸⁵ *Idem.*

Padrón de Postgrados de Excelencia de CONACYT, en las categorías de alto nivel o con reconocimiento internacional. En el CEDUA se imparte también la maestría en Estudios Urbanos y el Doctorado en Estudios de Población [...] Por otra parte, la maestría en Estudios de Asia y África y los doctorados en Economía y Estudios de Asia y África, se encuentran registrados en el Programa Integral de Fortalecimiento del Postgrado. En El Colegio de México también se imparten la maestría en Estudios de Género, las maestrías en Bibliotecología y Traducción, así como el doctorado en Estudios Urbanos y Ambientales.⁸⁶

Los posgrados que alberga son de alto nivel académico. Las estadísticas indican que:

En la actualidad el Colegio cuenta con tres centenares de alumnos de tiempo completo, la mayoría becarios en posgrados de excelencia, y otros tantos profesores-investigadores e investigadores de proyecto, todos sujetos a rigurosos mecanismos de evaluación académica. La proporción de estudiantes y profesores permite una relación estrecha entre ambos, conducente a la sólida formación de jóvenes investigadores.⁸⁷

En lo que a la investigación concierne, cuenta con distintas líneas de trabajo y con renombrados investigadores. Entre ellos destacan 18 profesores eméritos y un gran número de profesores incorporados al Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (CONACYT).

Una característica fundamental del COLMEX es la excelencia institucional. Cabe destacar que durante sus distintas administraciones ha habido acciones dirigidas a fortalecer tanto la calidad en la investigación como la educación de los alumnos. Así:

[...] desde sus orígenes en 1940 hasta ahora, esta pequeña institución especializada en ciertas áreas del conocimiento humanístico y de las ciencias sociales, es un ejemplo vivo de compromiso intelectual y académico con la creación y trasmisión de conocimientos en las diversas áreas de docencia e investigación que desarrolla. El prestigio ganado es un homenaje a las mujeres y hombres que desde los primeros pasos de La Casa de España hasta la actualidad han pasado por sus aulas enseñando y aprendiendo, y forjando como investigadores un importante corpus científico de primera importancia, testimonio de su compromiso y quehacer cotidianos.⁸⁸

La comunidad del COLMEX requiere información especializada para realizar sus labores, es por ello que “desde sus orígenes el Colegio se preocupó por tener una Biblioteca especializada.”⁸⁹ Ésta nació de la mano de la institución y en 1976 recibió el nombre de Biblioteca Daniel Cosío Villegas (BDCV).

La BDCV es una “biblioteca universitaria, especializada y de investigación, abierta a profesores, investigadores, alumnos, funcionarios de instituciones de educación superior y de otros sectores.”⁹⁰ Su objetivo es “apoyar los programas de

⁸⁶ *Idem.*

⁸⁷ *Idem.*

⁸⁸ *Idem.*

⁸⁹ *Idem.*

⁹⁰ El Colegio de México. *Portal de la Biblioteca Daniel Cosío Villegas* (México: El Colegio de México, 2010. (DE, 31 de Agosto, 2010: <http://biblioteca.colmex.mx>)).

investigación, docencia y difusión de El Colegio de México.”⁹¹ Para conseguirlo la BDCV realiza actividades tales como la selección, adquisición, conservación y canje de materiales, catalogación y clasificación de materiales. Ofrece además servicios de referencia y consulta, realiza préstamos internos, externos e inter-bibliotecarios, lleva a cabo un programa de formación de usuarios y establece relaciones de cooperación con otras unidades de información dirigidas a difundir y ampliar el acceso a los recursos de información.

Las colecciones de la biblioteca son sobre ciencias sociales y humanidades, áreas de interés para la institución. Se caracterizan por ser altamente especializadas en administración pública, economía, estudios internacionales, demografía, desarrollo urbano, historia, política, sociología, literatura, lingüística, traducción y estudios de la mujer. Se forman aproximadamente 700,000 volúmenes, equivalentes a 400,000 títulos de libros y folletos, 3,500 títulos de revistas y 2,000 discos compactos. Aunado a estos recursos, las colecciones integran un número significativo de base de datos también sobre ciencias sociales y humanidades.

La biblioteca está integrada por 18 bibliotecarios profesionales con especialización en bibliotecología o estudios de la información, dos profesionales de cómputo y 61 empleados administrativos.

Las instalaciones de la biblioteca son importantes, es por ello que el edificio donde se ubica fue diseñado específicamente para albergarla. Cuenta con espacios confortables para los usuarios, zonas idóneas para las colecciones, mostrador de préstamo, área de reproducción de documentos, guardarropa, sitios de lectura individual, cubículos para trabajo en grupo, sala de cómputo para estudiantes y un espacio para la consulta del catálogo.⁹² Durante sus 70 años de existencia ha ganado renombre tanto en el país como en otras regiones hasta el punto de ser considerada una biblioteca líder en México y en América Latina por la riqueza de sus colecciones, la calidad de sus procesos bibliográficos y la extensión de sus servicios a la comunidad académica nacional e internacional.

2.2. La indización en la BDCV

La BDCV lleva a cabo la indización de los documentos en sus instalaciones. Quienes realizan esta actividad son un grupo de profesionales en bibliotecología con estatus

⁹¹ *Idem.*

⁹² Además del acceso en sitio, el catálogo de la BDCV puede ser consultado vía Internet.

académico.⁹³ El proceso se desarrolla en la Coordinación de acceso a bases de datos, en donde trabajan cinco profesionales bajo la supervisión de un coordinador.

Cada catalogador realiza la catalogación completa de los materiales que le son asignados, es decir, lleva a cabo la catalogación descriptiva y la indización o catalogación por materias.⁹⁴ En la biblioteca, la catalogación por materias es el procedimiento mediante el cual se asignan encabezamientos de materia útiles para que los usuarios recuperen la información que buscan por temas.⁹⁵ Asimismo, es un proceso que se lleva a cabo en dos etapas. En la primera, denominada análisis temático de la obra:

Es necesario examinar la obra con todo detalle, si se desea realizar un trabajo eficiente en cuanto a asignación de encabezamientos de materia; en ocasiones es conveniente revisar el texto cuidadosamente y muchas veces leer trozos completos; sólo entonces el catalogador estará en situación de determinar cuál es en realidad el tema de la obra que tiene en sus manos.⁹⁶

En la segunda etapa, determinación de los temas, el catalogador elige el o los encabezamientos de materia adecuados al contenido del libro. La política a seguir al asignar los encabezamientos es: “uniformidad, consistencia, especificidad y conveniencia de uso”.⁹⁷ Esta actividad genera:

La palabra o palabras con que se define el contenido intelectual de una obra: incluye tópicos, personas, organizaciones y lugares. De esta manera, además de las entradas de autor, título, etc., las obras quedan representadas en el catálogo al público mediante los temas que tratan, esto permite a los usuarios saber qué obras posee la biblioteca sobre un asunto determinado.⁹⁸

En la biblioteca, la catalogación temática o indización es una actividad importante. La función de los encabezamientos de materia “los convierte en importantes

⁹³ En la biblioteca hay dos tipos de personal: administrativo y académico. El administrativo no tiene perfil profesional en bibliotecología y se encarga de brindar el servicio de préstamo en mostrador, el servicio de consulta, intercalar los materiales en estantería y apoyar actividades técnicas en las áreas de procesos técnicos y desarrollo de colecciones. El personal académico tiene perfil profesional en bibliotecología y se encarga de tareas más especializadas como son el servicio de referencia, catalogar y clasificar los materiales, coordinar el acceso a bases de datos, coordinar los servicios al público, coordinar actividades académicas vinculadas con los centros y realizar investigaciones sobre bibliotecología.

⁹⁴ En la BDCV es poco frecuente escuchar hablar de indización. Tanto en la descripción de su actividad como en los manuales que apoyan su trabajo se refieren a la actividad como “catalogación por materias”. Su concepto (Catalogación por materias) coincide con el de indización, establecido en el marco teórico, por lo que en este trabajo catalogación por temas, ce indización son usados como sinónimos.

⁹⁵ Arriola Navarrete, Oscar. *Documentos básicos para la catalogación de materiales en la Biblioteca Daniel Cosío Villegas* (México: Biblioteca Daniel Cosío Villegas, El Colegio de México, 1997), pp. 12-13.

⁹⁶ *Ibidem.* p. 14.

⁹⁷ Biblioteca Daniel Cosío Villegas, Coordinación de acceso a bases de datos. *Políticas de catalogación temática.* (México: Biblioteca Daniel Cosío Villegas, Coordinación de acceso a bases de datos, 2010. (DE, 5 de noviembre, 2010: <http://biblio.colmex.mx/Intranet/corabd.politicascatalog.html#autoridades>)).

⁹⁸ Oscar Arriola Navarrete. *Documentos básicos para...Op.cit.* p.36

auxiliares de los servicios que presta la biblioteca, puesto que permiten al usuario y al bibliotecario localizar en forma rápida y segura el material necesario para contestar preguntas sobre un tema dado.”⁹⁹

Para la catalogación por temas la biblioteca usa la lista de LCSH; sin embargo, como esta lista está en inglés y la biblioteca asigna sus temas en español, para la traducción de éstos se vale de otras herramientas que auxilian al programa de control de autoridades temáticas.¹⁰⁰ Sobre lo anterior, se enfatiza que: “Toda otra fuente, incluyendo: las listas de autoridades o encabezamientos de materia, ya sea impresas o en CD-ROM, en español o inglés, deberán considerarse únicamente como recursos de consulta.”¹⁰¹

2.3. La búsqueda por temas en el catálogo de la BDCV

La búsqueda por temas puede hacerse dentro de la biblioteca o consultando, vía Internet, el catálogo en línea. El acceso a los temas es posible de dos formas: en búsqueda por índices o en búsqueda por petición.¹⁰²

La interfaz de búsqueda en índices es la siguiente:

⁹⁹ *Idem.*

¹⁰⁰ Éstas son: United States. Library of Congress. *Bilindex : una lista bilingüe en español e inglés de encabezamientos de materia, equivalentes en español de los encabezamientos de la Biblioteca del Congreso de Estados Unidos de Norteamérica* (Oakland, Calif.: California Spanish Language Data Bada Oakland, 1984.); España. Unidad de Coordinación de Bibliotecas. *Lista de encabezamientos de materia de la Red de bibliotecas del CSIC* (Madrid: Consejo Superior de Investigaciones Científicas, 1995.); Biblioteca Luis Angel Arango. *Lista de encabezamientos de materia para bibliotecas* (Santafé de Bogotá: R. Eberhard, 1998.); fuentes de consulta sobre la materia; decisiones colegiadas entre catalogadores y consultas a los investigadores de El Colegio de México. Para la creación y/o validación de temas sobre México y América Latina también se auxilian de: Escamilla González, Gloria. *Lista de encabezamientos de materia* (México: UNAM, Instituto de Investigaciones Bibliográficas, 1978.). Mientras que para la creación y/o validación de temas sobre España y autores españoles consultan: España. Biblioteca Nacional. *Autoridades de la Biblioteca Nacional* (Madrid, Biblioteca Nacional, 1996.).

¹⁰¹ Biblioteca Daniel Cosío Villegas, Coordinación de acceso a bases de datos. *Políticas de catalogación...Op.cit.*

¹⁰² Esta tesis sólo indaga sobre las búsquedas por petición en las que el usuario formula la búsqueda sin ayudas para encontrar el tema que necesita.

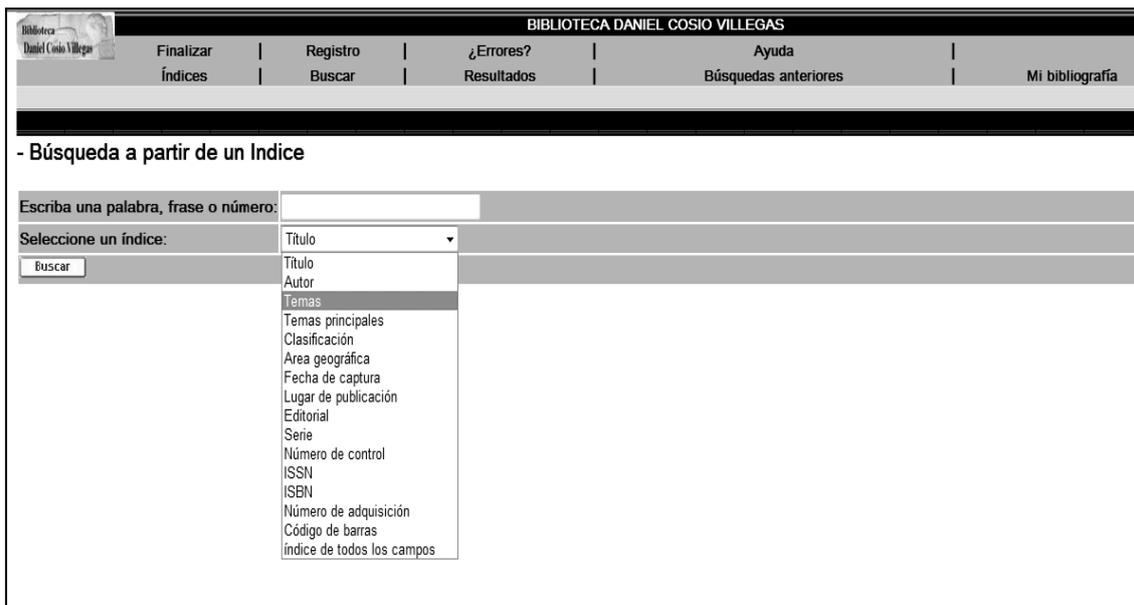


Fig.5. Interfaz de búsqueda en índices

La otra opción, mediante petición, se ve del siguiente modo:

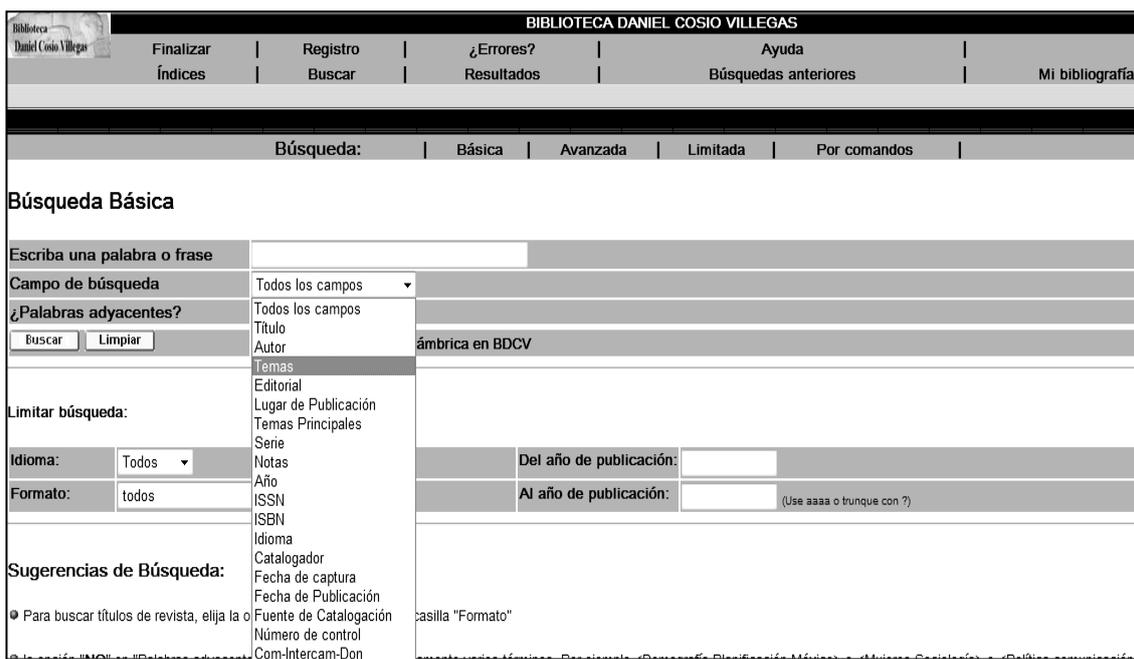


Fig. 6. Interfaz de búsqueda de la BDCV

Como puede observarse, hay una celda en donde el usuario introduce la palabra o palabras y del menú desplegable selecciona el campo de búsqueda.

La BDCV lleva a cabo investigaciones para mejorar sus actividades, no obstante, poco es lo que se ha indagado sobre el uso del catálogo. El único estudio al respecto consistió en una investigación sobre las búsquedas que no obtuvieron documentos. Su objetivo fue determinar patrones de búsquedas fallidas con la finalidad de diseñar un

sistema de ayudas que le permitieran al usuario construir búsquedas exitosas. Aunque el artículo trata problemas de todas las opciones que el catálogo proporciona, se detectaron algunos vinculados con la exploración por temas. Éstos serían:

- a) Naturaleza de la colección: búsquedas que por su temática no correspondían con el contenido de las colecciones.
- b) Diseño del sistema: son producidos porque el sistema no reconoce caracteres que el usuario introduce, por ejemplo, signos de puntuación o errores ortográficos.
- c) Sintaxis propia de los buscadores de Internet
- d) Peticiones en lenguaje natural que no pueden ser enlazadas con la indización.

Resulta un trabajo valioso, ya que detecta los errores más frecuentes que ocasionan búsquedas fallidas y, al mismo tiempo, posibilita el establecimiento de líneas de investigación más profundas sobre los problemas detectados. Debido a que es el único trabajo sobre el asunto expone la necesidad de una mayor investigación al respecto.

2.4. El lenguaje de la indización y de la búsqueda por temas en la BDCV

La indización de documentos en la BDCV es monitoreado por medio del control de autoridades temáticas. El proceso se auxilia de lenguajes artificiales de donde se seleccionan las palabras útiles para representar el contenido de los documentos. Durante muchos años han optado por usar la lista de encabezamientos de LCSH como herramienta. El motivo de ello es por:

La relevancia del uso de LCSH es que su representación y organización del conocimiento es la más completa y extensa; está actualizada, es fiable y de excelente calidad en cada registro. Lo que le sustenta el por qué es usada como modelo no sólo en Estados Unidos, sino en muchas otras ciudades. Además, tiene otra ventaja, que está disponible en formato electrónico.¹⁰³

Al emplear esta herramienta se han enfrentado a algunos problemas, los más relevantes son el idioma inglés de la lista y el contexto cultural al que corresponde. Debido a eso:

¹⁰³ Reynaldo D. Figueroa-Servín y Berta Enciso. "Subject authority control at El Colegio de Mexico's library: the whats and hows of a project" en *Cataloging & classification quarterly* (vol. 32, núm. 1, 2001), pp. 65-80.

El equipo de control de autoridades de la BDCV emplea traducciones y representaciones de los temas en español, siempre siguiendo la estructura de LCSH; sin embargo, algunas veces se le da un tratamiento diferente a los temas en relación con nuestra cultura, incluyendo cultural latinoamericanas y del mundo hispánico en general.¹⁰⁴

Otro trabajo, *Lenguajes de indización libres o controlados: la experiencia de la Biblioteca Daniel Cosío Villegas de El Colegio de México*¹⁰⁵, señala que: “la cuestión de qué tipo de lenguaje de indización adoptar presenta mayor interés si tenemos en cuenta que, como han demostrado numerosas encuestas en los Estados Unidos, al menos la mitad de las búsquedas en los catálogos en línea son temáticas.”¹⁰⁶ La autora agrega que, en su opinión, el principal beneficio que puede reportar el uso de un lenguaje de indización controlado es la eficacia en la recuperación. En este sentido, su punto de vista coincide con la teoría sobre lenguaje de indización cuando dice que el lenguaje natural, por naturaleza, presenta problemas de sinonimia y polisemia que repercuten en la obtención de información.

En la BDCV el uso de un lenguaje controlado implica “la presencia de personal especializado en análisis documental.”¹⁰⁷ Aunado a esto, el personal debe tener el conocimiento experto necesario para el trabajo del control de autoridades, esto es, crear registros: personales, corporativos, de reuniones y conferencias, autoridades de título y autoridades de materia, como parte del control bibliográfico.

La BDCV ha puesto mucho énfasis para lograr la consistencia en el uso del lenguaje de indización, pues, las inconsistencias en éste, pueden dar lugar a la dispersión de documentos sobre un mismo asunto. Al respecto Moreno dice que si los documentos que tratan, por ejemplo, sobre “Taquiografía” tienen asignados indistintamente los temas “Taquiografía” y “Estenografía” al buscar por la palabra “Taquiografía” se recupera únicamente los documentos que tenía asignado ese tema y no los que tenían como tema “Estenografía”. Por lo que resulta vital aplicar principios de consistencia en la catalogación temática y crear vínculos entre sinónimos o palabras relacionadas.

¹⁰⁴ *Idem.*

¹⁰⁵ Pilar Moreno Jiménez. “Lenguajes de indización libres o controlados: la experiencia de la Biblioteca Daniel Cosío Villegas de El Colegio de México”. *XII Coloquio sobre Investigaciones Bibliotecológicas* (México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 7-8 de Julio de 1994), Mesa redonda Lenguajes de Indización Libres o Controlados.

¹⁰⁶ *Ibidem.* p. 1.

¹⁰⁷ *Ibidem.* p. 6.

Ante la necesidad de los usuarios por recuperar la información y el interés de la BDCV para facilitarlos, la biblioteca, lleva a cabo un Programa de Control de Autoridades de Materia,¹⁰⁸ pues considera que: “Le permiten al usuario recuperar los datos deseados no sólo a través de las formas “autorizadas” o aceptadas, sino también a partir de las referencias cruzadas, que pueden ser las “canceladas” o no autorizadas y las relacionadas a estas últimas.”¹⁰⁹

El control de autoridades busca apoyar el uso de los lenguajes controlados direccionando las solicitudes de los usuarios expresadas en formas no autorizadas (lenguaje natural) hacia las formas autorizadas (lenguaje controlado). Al respecto, el proyecto de autoridades enfatiza que:

Las bibliotecas o centros de información que han introducido los catálogos de autoridad en sus entidades apoyan la idea de que se requiere un lenguaje controlado que aunque no universal, esté lo más normalizado posible. Sin embargo, la idea de universalidad es una mera utopía, principalmente porque es casi imposible cubrir todos los aspectos de un tema determinado.¹¹⁰

Ya se ha hablado acerca del lenguaje de indización en la biblioteca. Resulta ahora indispensable tratar el lenguaje de exploración. Los usuarios de la BDCV son expertos en ciencias sociales y humanidades y cuentan con una formación académica sólida. En su mayoría conocen la terminología de la disciplina en que investigan; no obstante, su conocimiento está dado en grados distintos. Hay usuarios muy expertos (estudiantes de doctorado e investigadores), expertos (estudiantes de maestría), semiexpertos (estudiantes de licenciatura) y novatos (aquellos que apenas entran en contacto con la disciplina de estudio).

Todos ellos en sus peticiones emplean el lenguaje natural caracterizado por la pluralidad de significados y la variedad de palabras para nombrar un referente. Se infiere que en sus búsquedas por temas emplean palabras de su vocabulario (léxico disponible), enriquecido por la literatura del área de estudio, sus profesores y los eventos (conferencias, congresos, encuentros) en los que participan. Su lenguaje es el de

¹⁰⁸ El programa se denomina Programa cooperativo de Control de Autoridades y en él participan varias bibliotecas académicas de México. Para mayores referencias véase: Figueroa Servín, Reynaldo D. y Berta Enciso. “Subject authority control at El Colegio de Mexico's library: the whats and hows of a project” en *Cataloging & classification quarterly*. Nueva York: Haworth, vol.32, núm.1, 2001), pp. 65-80; Quijano-Solis, Alvaro, Pilar Maria Moreno-Jimenez y Reynaldo D. Figueroa-Servin. “Automated authority files of Spanish-language subject headings” en *Cataloging & classification quarterly* (Nueva York: Haworth, vol. 29, núm. 1-2, 2000), pp.209-223 y Programa Cooperativo de Control de Autoridades (PCCA), Comité Técnico. *Políticas para la elaboración de autoridades de materia* (México: El Comité, 2009.).

¹⁰⁹ Programa Cooperativo de Control de Autoridades (PCCA), Comité Técnico. *Políticas para la elaboración de autoridades de materia* (México: El Comité, 2009), p. 7.

¹¹⁰ *Idem*.

un profesional especializado o en proceso de especializarse en ciencias sociales o humanidades. Estamos, entonces, ante un grupo heterogéneo de usuarios cuyo manejo de la terminología de la disciplina y habilidades informativas se desconocen, pero influirán directamente en cómo buscan por temas y el lenguaje con el que buscan.

2.5. Conclusión

La BDCV es una biblioteca que debe responder a las altas exigencias de los usuarios de El Colegio de México, quienes constantemente requieren información para realizar sus labores académicas, docentes o de investigación. Debido al contexto donde se ubica, caracterizado por la excelencia institucional, la biblioteca busca mantenerse a la vanguardia tanto en el desarrollo de sus actividades como en su oferta de servicios. Durante sus 70 años de historia ha logrado posicionarse como una biblioteca líder en ciencias sociales y humanidades en el ámbito latinoamericano.

La indización o catalogación temática realizada en la BDCV es muy apegada a la teoría de la actividad. En tanto que sus objetivos son servir para la organización, localización y recuperación de los materiales y se constituye de dos etapas: en la primera se analiza el documento para determinar de qué trata (*aboutness*) y en la segunda se asignan los términos de indización. Aunado a ello, el trabajo de indización está normado por un meritorio trabajo de control de autoridades temáticas que pocas bibliotecas mexicanas, aún las académicas, realizan. La biblioteca invierte muchos recursos en la organización de la información, pero ha dejado de lado la otra cara de la moneda: la recuperación. A corto plazo, se vislumbran necesarias investigaciones para la retroalimentación y mejora de sus actividades, específicamente en el contexto de la recuperación temática.

Capítulo 3: Fundamentos metodológicos

3.1. Instrumentos de análisis

El objetivo de este trabajo es determinar las implicaciones del encuentro entre el lenguaje del usuario y el lenguaje de indización a partir de la recuperación temática en el catálogo electrónico de la BDCV. Para alcanzarlo, esta tesis usa dos instrumentos de análisis: transacciones de búsquedas por temas y cuestionario.

3.1.1. Transacciones de búsquedas por temas

3.1.1.1. Justificación

El análisis de transacciones es un método reciente para la recolección de datos. Es posible gracias a que los catálogos cuentan con un sistema que permite llevar el control estadístico de todas las acciones registradas en él. En este trabajo se decidió optar por este instrumento puesto que en las peticiones de búsqueda se observa la formulación que el usuario hace al sistema en lenguaje natural y la respuesta que el sistema ofrece luego de enfrentar su petición a la indización hecha en lenguaje controlado. Así, en las transacciones por temas es posible observar las interacciones entre ambos lenguajes mediante una respuesta satisfactoria o negativa del sistema expresada en términos de recuperación de información.

Además de la razón anterior, hubo dos motivos más para optar por las transacciones de búsquedas. El primero fue su carácter no intrusivo. Esto significa que los datos obtenidos son producidos por el usuario sin que haya un observador que modifique su conducta. El segundo indica que en México no hay investigaciones que empleen las transacciones como instrumento por lo que vale la pena indagar sobre las ventajas de su utilización.

3.1.1.2. Acopio de datos

El instrumento exige tomar en cuenta tres aspectos antes de realizar el acopio de los datos: los usuarios de quienes se obtendrán las transacciones, el número de transacciones y el periodo en el que éstas serán recolectadas. En la tesis, el tratamiento de los aspectos establecidos quedó como sigue:

- 1) Los usuarios que produjeron las transacciones fueron usuarios finales del sistema.

2) El tamaño de la muestra debía oscilar entre 600 y 700 transacciones, por tres motivos:

- Representatividad: una muestra de este tamaño resulta representativa en tanto que, como señala Rumi Y. Graham¹¹¹, el tamaño de la muestra puede introducir sólo pequeñas diferencias en los resultados finales.
- Adecuación: una muestra entre 600 y 700 transacciones resulta ser la cifra media usada en investigaciones similares.
- Tiempo de análisis: otro elemento que se tomó en cuenta fue el tiempo que implicaba el análisis de los registros. Según los tiempos establecidos para el término de la investigación una muestra con este número de transacciones era idónea en función de los tiempos fijados.

3) El periodo de recolección de las transacciones fue de una semana.¹¹² Abarca del 30 de agosto al 4 de septiembre de 2010. Los estudios similares resaltan la importancia de tomar la muestra en una fecha en la que se refleje el “comportamiento promedio” de la comunidad a la que sirve la biblioteca, es decir, no tomarla al principio del semestre, periodos vacacionales, semanas que incluyan asuetos, etc. Este aspecto se tomó en cuenta, ya que esta semana forma parte del semestre ordinario de la institución y cumple con el requisito enunciado.

La obtención de la muestra fue posible gracias al Sistema Integral de Gestión de Bibliotecas (SIGB) con el que cuenta la BDCV (Aleph). Éste tiene una base de administración interna denominada Vir-01 donde se registran todas las operaciones llevadas a cabo durante cierto periodo de tiempo y en la cual resulta posible la manipulación de la información. Los datos almacenados en la bitácora pueden

¹¹¹ Graham, Rumi Y. “Subject no-hits searches in an academic library online catalog: an exploration of two potential ameliorations” en *College and research libraries* (Chicago: American Library Association, vol. 65, num.1, 2004), pp. 36-54.

¹¹² Las investigaciones que usan transacciones en catálogo establecen distintos tiempos para recopilar la muestra. Los periodos van desde las transacciones de un día (69) hasta las producidas en 120 días (1524). En este trabajo se estableció una semana, porque fue el tiempo estimado para recoger entre 600 y 700, que fueron fijadas como tamaño de la muestra. Para profundizar más sobre las muestras en análisis de transacciones temáticas de catálogos véase Graham, Rumi Y. “Subject No-hits Searches...*Op.cit.*”

manipularse de acuerdo a campos bibliográficos establecidos (autor, título, tema, ISBN, etc.).

De esta bitácora, mediante un algoritmo de filtrado para formato de texto, se seleccionaron todas aquellas búsquedas por tema y se enviaron a un documento nuevo.¹¹³ Una vez en este documento, se eliminaron columnas no útiles que registraban datos como la hora o fecha de la transacción y sólo se dejaron las 660 transacciones producidas en una semana que forman la muestra.¹¹⁴

3.1.1.3. Preparación de datos

Aunque la muestra ya estaba filtrada, fue necesario llevar a cabo un tratamiento previo que consistió en:

- 1) Eliminación de palabras vacías¹¹⁵ (y, de, la, en, del, e, el, a, of, in, and, their, es, su, una, lo, the, sobre, un, antes, según, sus) que el usuario enviaba como peticiones de búsqueda. Se asume que fueron errores probablemente debidos a envíos imprevistos.
- 2) Eliminación de registros duplicados que el sistema por alguna razón, no determinaba, repetía.

3.1.1.4. Análisis de datos

Para el análisis de la muestra se crearon categorías en las que se pudieran agrupar las transacciones¹¹⁶. Se crearon cuatro categorías de análisis con sus correspondientes sub-categorías:

¹¹³ En la etapa de obtención del corpus resulta importante señalar la labor del Ing. Francisco Mejía, Coordinador de Cómputo de la BDCV, quien llevó a cabo las operaciones necesarias para contar con la muestra.

¹¹⁴ La muestra de transacciones temáticas se incorpora en la sección de anexos: *Anexo I*.

¹¹⁵ Hay dos tipos de palabras: palabras plenas que tienen un referente ya sea real o imaginario (libro, cama, amor, soledad) y palabras vacías que no tiene un referente, son artículos, preposiciones, conjunciones, etc. (un, una, los, las).

¹¹⁶ Algunas de las categorías fueron retomadas de investigaciones similares, especialmente de dos: Lester, Marilyn Ann. *Coincidence of user...Op.cit.* y de Graham, Rumi Y. "Subject No-hits Searches..." *Op.cit.* Sin embargo, según lo requirió la investigación, se fueron agregando otras clases, porque había casos que no tenían cabida en las categorías establecidas en las investigaciones mencionadas.

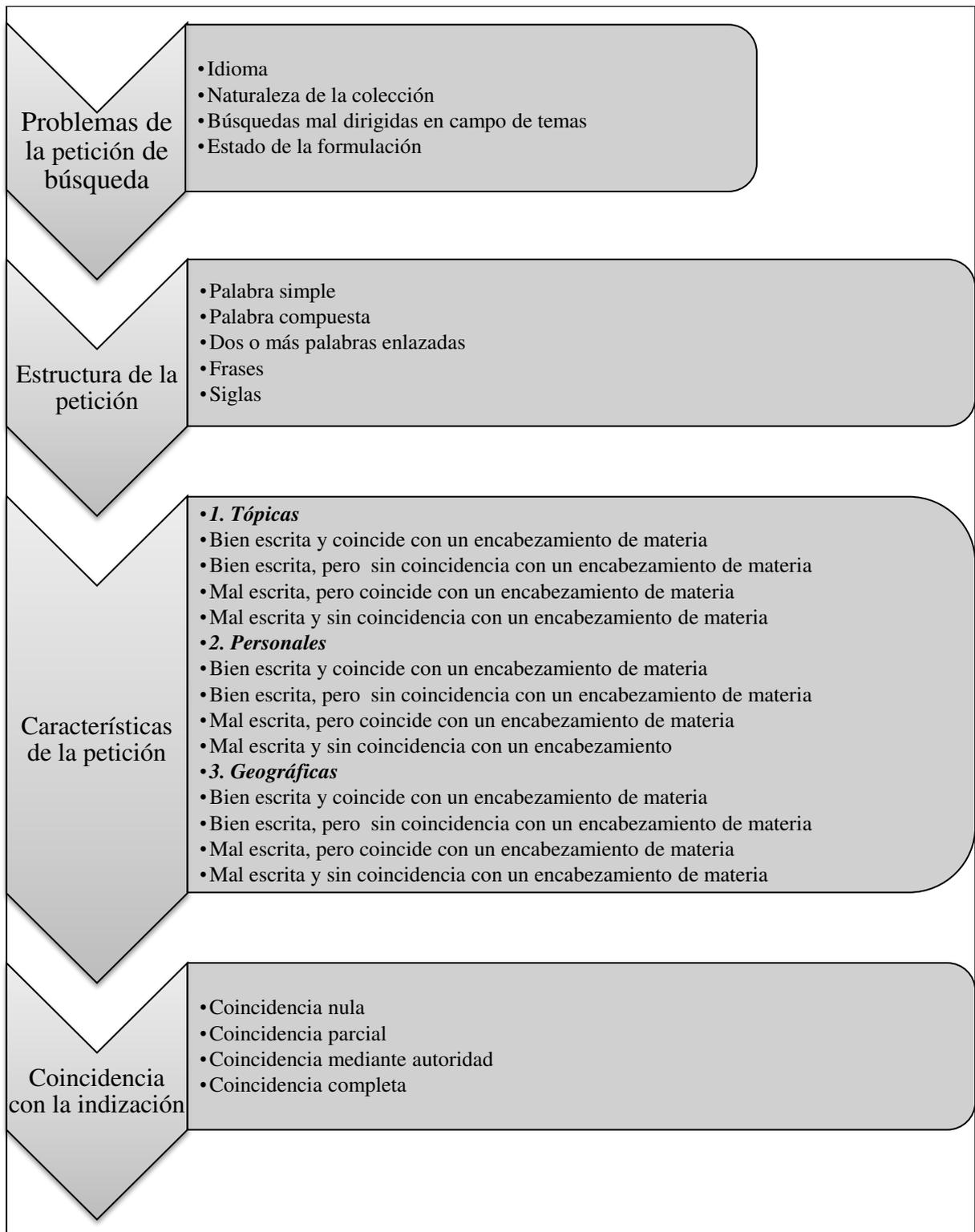


Fig. 7. Categorías de análisis de las transacciones temáticas

Para determinar las implicaciones del encuentro entre el lenguaje del usuario y el lenguaje de indización fue necesario crear las categorías anteriores bajo las cuales la información pudiera ser analizada. Se trabajó primero en el análisis de la petición de búsqueda lo que justifica las categorías: “Problemas de la petición de búsqueda” y

“Estructura de la petición”. Una vez definidos estos aspectos, se encontró que había otros rasgos de la petición que influían en su enlace con el lenguaje de indización. Estos rasgos de la petición dieron origen a la segunda categoría de análisis denominada “Características de la petición”. Finalmente, la tercera categoría, “Coincidencia con el lenguaje de indización”, indaga específicamente sobre la función de coincidencia entre el lenguaje natural y el lenguaje controlado que influye en la recuperación de información.

3.1.1.5. Presentación de resultados

Los resultados obtenidos en las transacciones temáticas se presentan en apartados, según las cuatro categorías generales establecidas para el análisis. Los tres primeros apartados tratan sobre la petición de búsqueda y el cuarto sobre la coincidencia de ésta con la indización. En cada apartado se presentan gráficos que explican la presencia de las categorías en el corpus trabajado y también se brindan ejemplos representativos de las categorías.

3.1.2. Cuestionario

3.1.2.1. Justificación

El análisis de transacciones es un buen método para la recolección de datos, pero presenta ciertas desventajas. Una de las más evidentes es que deja escapar muchos eventos o impresiones que intervienen en los datos que registra, esto es, no ofrece información sobre la experiencia del usuario en su interacción con el sistema ni tampoco sobre las intenciones de los usuarios y sus necesidades. Wildemuth¹¹⁷ dice que la única manera de superar estas limitaciones es combinar las transacciones con otro método de obtención de datos.

En esta tesis el enfoque está centrado en el usuario, pues se intenta saber las implicaciones que tienen sus peticiones expresadas en lengua natural al enfrentarse a una indización hecha en lenguajes controlados. Las transacciones de búsqueda muestran este aspecto; sin embargo, para fines de esta investigación interesa también saber qué opina el usuario de este hecho, su experiencia en la búsqueda temática y su opinión sobre los materiales que recupera por temas. Por eso, se optó por un segundo

¹¹⁷ Sheble, Lallura y Barbara M. Wildemuth. “Transactions logs” en Bárbara M. Wildemuth, ed. *Applications of social research methods to questions in information and library science* (Westport, Conn: Libraries Unlimited, 2009), p.168.

instrumento, cuestionario,¹¹⁸ con el que se obtuvo la impresión de los usuarios sobre tales aspectos.

3.1.2.2. Diseño

El diseño del cuestionario se hizo en tres etapas. En la primera se construyeron las categorías con sus correspondientes preguntas, la segunda, correspondió al piloteo del instrumento, y en la tercera, se hicieron modificaciones al cuestionario en función del piloteo.

El cuestionario final¹¹⁹ quedó constituido por 22 preguntas agrupadas bajo 7 categorías que guardan relación con:

- a) El uso de la búsqueda temática
- b) La opinión de los usuarios sobre la búsqueda
- c) La formulación de búsqueda
- d) La petición de búsqueda
- e) Conocimientos de los usuarios sobre los lenguaje controlados
- f) La recuperación en la búsqueda
- g) Sugerencias de los usuarios para mejorar la búsqueda

La primera categoría trata sobre el uso de la opción. Puesto que esta es la primera investigación formal llevada a cabo en la BDCV sobre el tema, resulta obligado saber si usan o no el campo de búsqueda. Las preguntas 1 y 2 forman parte de esta categoría.

El segundo grupo de preguntas se centra en la búsqueda temática. Comprende las preguntas 3, 4, 5, 6 y 7 que están enfocadas a conocer la opinión de los usuarios. Como ejemplo, el deseo de saber si la búsqueda por materias les parece difícil o fácil, los motivos por los que la perciben de este modo y qué parte del proceso de la búsqueda temática les resulta más complicado.

La tercera categoría cuestiona sobre la formulación o interrogación del usuario. Como se señaló en el marco teórico, la búsqueda temática empieza con una conceptualización o descripción de tema que se quiere, luego le sigue un momento de

¹¹⁸ Frente a la opción de aplicar entrevistas, cuestionarios o hacer un grupo focal se prefirió el cuestionario, porque la entrevista está más dirigida a aspectos cualitativos como historias de vida o experiencias y el grupo focal recoge impresiones grupales. Por su parte, el cuestionario resultó ser un instrumento más eficaz para recopilar puntos de vista individuales y sobre aspectos concretos que era lo que en este trabajo se necesitaba.

¹¹⁹ El cuestionario aplicado se incorpora en la sección de anexos: *Anexo 2*.

formulación frente al sistema de recuperación de información que la teoría define como un momento álgido. De este modo, resulta ineludible saber si para los usuarios este momento es complicado. Las preguntas de la categoría son las 8, 9, 10, 20 y 21.

La cuarta categoría, formada sólo por la pregunta 14 pretende saber cómo son, en función de estructura, las peticiones del usuario. Se incluyó porque muchas investigaciones que tratan sobre la búsqueda temática, advierten que la búsqueda por materias está tomando una forma más bien de búsqueda por palabras clave. Esta pregunta en combinación con los resultados del análisis de transacciones servirá para afirmar o desmentir tal aseveración.

La quinta categoría indaga acerca del modo en que el usuario conoce los temas que la biblioteca asigna a los materiales y los revisa. Es decir, si entra en contacto con el lenguaje controlado y si sabe de su existencia en los registros. En este grupo se insertan las preguntas 11,12, y 13.

La sexta categoría del cuestionario trata sobre la recuperación de información a partir de la petición de búsqueda. Se divide en dos situaciones: cuando el usuario recupera documentos, preguntas 15 y 16; y cuando no recupera, preguntas 17,18 y 19. Cuando recuperan información se les preguntó sobre la utilidad de los documentos recuperados y la relación de éstos con el tema que tenían en mente. Cuando no recuperan documentos, se les preguntó por qué creían que se había presentado esta situación y qué habían hecho ante tales resultados. Los usuarios manifestaron que se han encontrado en ambas situaciones.

La última categoría que sólo contiene una pregunta, la número 22, es abierta y brinda al usuario la posibilidad de hacer sugerencias sobre las acciones que la biblioteca deberá considerar para que la búsqueda por temas ofrezca mejores resultados.

3.1.2.3. Acopio de datos

El cuestionario se aplicó a 50 usuarios finales del sistema: estudiantes, investigadores, becarios y usuarios externos. En total fueron 30 estudiantes de los distintos programas de estudio que ofrece la institución; 5 investigadores; 5 becarios y 10 usuarios externos. Todos los encuestados respondieron el cuestionario en formato impreso en un tiempo aproximado de 16 minutos.

3.1.2.4. Análisis de datos

El análisis de las respuestas de los cuestionarios se llevó a cabo mediante el programa *Excel* de *Microsoft Office*. Las respuestas de cada pregunta se contabilizaron, siempre que fue posible, de mayor a menor frecuencia. En función de la frecuencia de las respuestas obtenidas se hizo el análisis de la información. En las preguntas abiertas que presenta el cuestionario, se crearon categorías de estudio de acuerdo a las respuestas obtenidas.

3.1.2.5. Presentación de resultados

Los resultados obtenidos del cuestionario se presentan en apartados que corresponden a las siete categorías bajo las cuales fueron agrupadas las preguntas. Igual que en la presentación de resultados de las transacciones temáticas se insertan gráficos donde se muestran las preguntas con sus porcentajes de respuestas. En algunos casos se insertan citas textuales, extraídas de las preguntas abiertas, que se consideró eran destacables.

3.2. Conclusión

En México no hay investigaciones que indaguen sobre la búsqueda por temas en catálogos de bibliotecas y menos aún sobre el encuentro entre el lenguaje de indización y el lenguaje de la búsqueda en la recuperación por materias. La literatura dominante al respecto es norteamericana y en ella destacan autores como Ray Larson, Karen Markey, Macia Bates y Jennifer Rowley quienes han llevado la investigación al terreno práctico. Las metodologías empleadas por los autores en sus trabajos son el punto de partida obligatorio para quienes estén interesados sobre el tema.

La metodología empleada en esta investigación se nutre de los trabajos de estos autores; sin embargo, en función de los objetivos específicos y el enfoque dominante dirigido al usuario, se incorporaron elementos que poco a poco dieron origen a una metodología propia. Para el logro de sus objetivos, la tesis se vale de dos instrumentos:

a) Transacciones de búsquedas por temas: donde es posible observar el encuentro entre la petición del usuario y la respuesta del sistema a partir de la confrontación de la solicitud con la indización.

b) Cuestionario: donde se recoge la impresión de usuario respecto a su experiencia en la búsqueda por temas, la recuperación en ésta y su impresión sobre la función de enlace entre su petición y la respuesta del catálogo.

Las metodologías usadas buscan complementarse ya que en las transacciones se obtuvieron datos precisos sobre los fenómenos que se presentan en el encuentro entre ambos lenguajes a través de las transacciones y en el cuestionario se completa el panorama integrando el punto de vista del usuario que no es posible de observar en las primeras. La experiencia de usar estas metodologías fue satisfactoria, pues se obtuvieron buenos resultados y se corroboró que el uso de varios instrumentos dirigidos al logro de un objetivo brinda un panorama más amplio del problema de investigación.

Capítulo 4. Resultados obtenidos mediante las transacciones de búsqueda

4. 1. La petición de búsqueda

4.1.1. Problemas frecuentes

4.1.1.1. Idioma

La BDCV se caracteriza por tener una de las colecciones más completas sobre ciencias sociales y humanidades en América latina. Es bien sabido que, además de tener materiales en idioma español, gran parte de su colección está en otros idiomas. Estos documentos responden a la necesidad de los usuarios de contar con obras que integren distintas visiones sobre los temas. Así, la biblioteca integra obras con contenidos en inglés, francés, alemán, italiano, ruso, lenguas de Medio oriente, Asia y África.

Es común que los usuarios consulten materiales en lenguas distintas al español, pues en su mayoría leen o hablan otros idiomas. Aunado a este hecho, el nivel de la institución exige tanto a investigadores como a estudiantes consultar documentos en otras lenguas, ya que gran parte de la bibliografía con más autoridad en ciencias sociales y humanidades está desarrollada por sectores de investigación internacionales, cuyas lenguas son distintas al español.

En el proceso de catalogación de documentos en otras lenguas, la BDCV sigue los principios de las *Reglas de Catalogación Angloamericanas*¹²⁰, conservando el título y otros elementos en el idioma original. En tanto que la política de indización establece asignar los temas en español. Sólo se agregan términos en inglés de la lista de LSCH a las autoridades temáticas y pueden agregarse términos en idiomas distintos como términos no autorizados bajo registros de autoridades.

Los usuarios de la biblioteca, en la búsqueda, emplean palabras en idiomas distintos al español. Por lo que se infiere que hay quienes desconocen que la biblioteca asigna los temas en español, quizá, mediante sus peticiones en otro idioma pretenden recuperar documentos cuyo contenido esté en el idioma con el que buscan, sin saber que la petición en español recupera obras en todas las lenguas.

Las búsquedas en otro idioma son frecuentes, obsérvese su presencia:

¹²⁰ *Reglas de Catalogación Angloamericanas: actualización de 2003* (Preparadas bajo la dirección del Joint Steering Committee for Revision of AACR. Traducción y revisión general de Margarita Amaya de Heredia. Bogotá: Rojas Eberhard, 2004.).

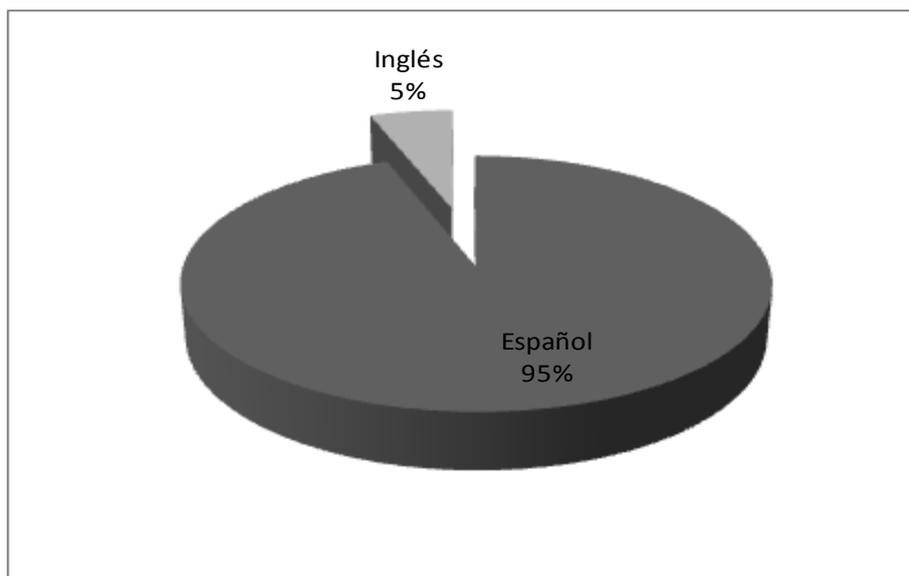


Fig. 8. Idioma de las peticiones de búsqueda

Es necesario recalcar que en la muestra analizada las peticiones en otro idioma sólo se registran en inglés y representan el 5% de la muestra. Ante estas búsquedas, surge el interés por saber si los usuarios obtuvieron documentos o no. Se infiere por tanto que no hubo recuperación, pues la indización se hace en español; aunque contra toda predicción se encontró que muchas de éstas extrajeron documentos. La recuperación de las consultas obtuvo los siguientes resultados:

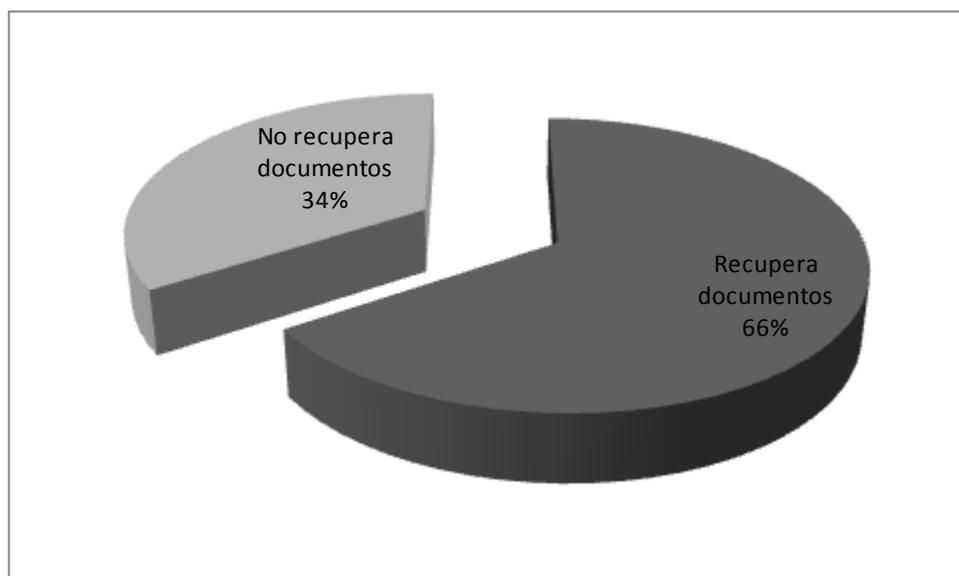


Fig. 9. Recuperación de peticiones en inglés

El 66% de estas peticiones consigue documentos gracias a que el término solicitado en inglés aparece como parte de un registro de autoridad. Tal es el caso de

consultas como “History of brazil”,¹²¹“Criminal law”, “Landscape”, “Sociologists” “Germany”, “Management”. Si bien estas consultas recuperaron documentos, resulta evidente que la simple agregación del término aceptado por LCSH a las autoridades temáticas de la biblioteca no resultará suficiente para dar respuesta a búsquedas expresadas en idioma distinto a español. Incluso, resulta fortuita la recuperación, ya que LCSH, herramienta auxiliar en la creación de autoridades, coincide con el idioma en que fueron plateadas las consultas.

De las peticiones en inglés, el 34% no tuvo recuperación. Debido a que fueron búsquedas más complejas como “Canada political system” “Marine biodiversity conservation” o “Democracy in the caribbean” que no coinciden con un encabezamiento aceptado por LCSH y, por tanto, no se registran en algún registro de autoridad.

El idioma en el cual es planteada la búsqueda es uno de los problemas más frecuentes de las peticiones. Aunque gran cantidad de las solicitudes obtuvo resultados, ésta es azarosa, por lo que debe tomarse una decisión ante el problema. Se vislumbran dos alternativas: indizar también en idioma inglés o instruir a los usuarios sobre el idioma de la indización. Lo primero se torna una tarea ardua, pues la BDCV ya destina gran cantidad de recursos a la indización y la posibilidad de asignar temas en idioma inglés se vislumbraría como un proyecto que implica mucho trabajo y pocos beneficios, en virtud de la ocurrencia del problema (5%). Ante tal panorama, la segunda opción, fortalecer la instrucción de los usuarios se vislumbra como la solución idónea. A la par de esta actividad, incluir en la autoridad en español su correspondencia en inglés parece ser una medida efectiva, pues ayudó para que el 66% de las peticiones consiguieran materiales.

4.1.1.2. Naturaleza de la colección

La colección de la BDCV es altamente especializada en ciencias sociales y humanidades, pues su objetivo principal es apoyar los programas docentes y las investigaciones desarrolladas en El Colegio de México. Sin embargo, en menor proporción cuenta con algunos materiales distintos a estas áreas.

¹²¹ Las peticiones de búsqueda empleadas para ejemplificar las categorías de análisis se colocan entre comillas para diferenciarlas del texto. Con fines de presentación del trabajo, no se presenta la petición como la introdujo el usuario, dado que muchas veces éste usa sólo letras mayúsculas o no coloca acentos. Para uniformarlas, siempre la primera letra es puesta en mayúscula y se agregan los acentos correspondientes. Si se desea ver, con exactitud, las palabras introducidas, éstas se puede consultar en la sección Anexos: Anexo 1, cuyo orden es alfabético.

La comunidad de usuarios internos de la BDCV tiene conocimiento sobre las temáticas que abarca la colección. Una de las prioridades de la biblioteca es especializar su colección de acuerdo con las disciplinas temáticas de interés, por ello, se han establecido principios muy claros de desarrollo de colecciones, Así, cuando los usuarios internos requieren materiales de áreas de conocimiento distintas a las ciencias sociales y humanidades, casos excepcionales, se da prioridad al préstamo interbibliotecario sobre la compra.

Los usuarios internos de El Colegio de México conocen la cobertura temática de la colección, empero, los usuarios externos no siempre lo saben. Otro de los problemas de la muestra es la presencia de peticiones con naturaleza ajena a la colección que no recuperan documentos. Asegurar que estas búsquedas fueron producidas por usuarios externos resulta arriesgado; sin embargo, es seguro mencionar que no fueron realizadas por usuarios asiduos a la biblioteca, pues ellos conocen la cobertura temática del material.

Las consultas de naturaleza distinta a las ciencias sociales y humanidades se presentan como sigue:

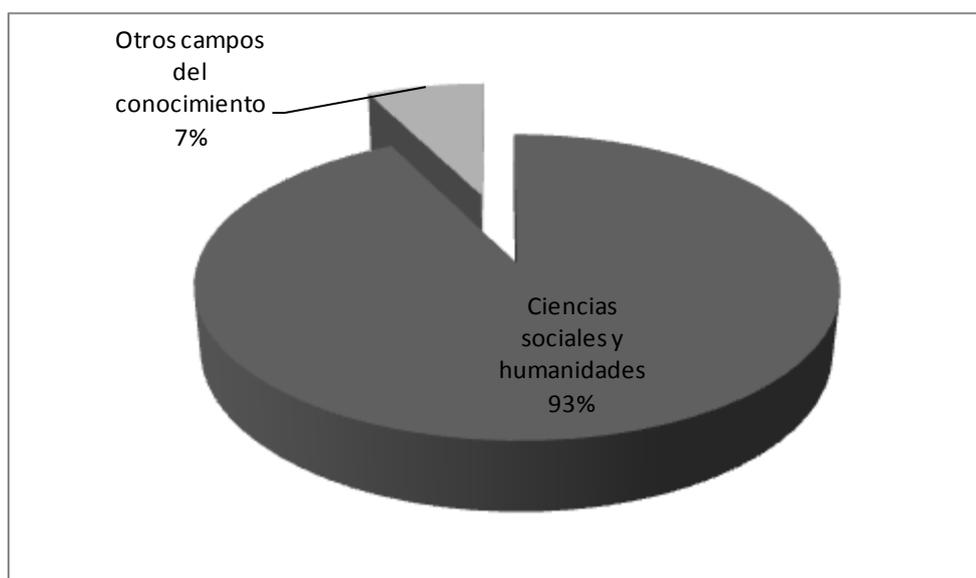


Fig. 10. Campo temático de las peticiones

Su presencia es del 7% y son búsquedas sobre Física, Medicina, Química, Biología, Cocina, Mecatrónica, etc. Se determinó que son peticiones bastante especializadas probablemente producidas por usuarios más o menos expertos en las disciplinas. Algunas de las peticiones que presenta la muestra son: “Clonazepam”,

“Aborto espontáneo”, “Homogenización”, “Centrifugación”, “Método de cross” y “Relación interespecífica de lacerillos”.

4.1.1.3. Búsquedas mal dirigidas en el campo temas

Las búsquedas mal dirigidas en campo de tema representan otro inconveniente. Pero, en comparación con los problemas del idioma o la naturaleza de la colección, su frecuencia es baja:



Fig. 11. Adecuación de la petición al campo temas

Representan sólo el 1% de la muestra. Lo más probable es que se deban a descuidos de los usuarios al seleccionar la opción de búsqueda. Son consultas que debieron ser dirigidas mediante la opción de autor, título o todos los campos. Así, “Alvaro Matute” debió ser una petición por autor, mientras que “Baudelaire y el artista de la vida”, “El último de los mohicanos”, “La araucana” y “Should sociologists forget their mothers and fathers” debieron ser búsquedas por título. Mientras que la petición “Baudelaire y el artista de la vida and azua” debió ser dirigida mediante la opción todos los campos.

En función de la baja presencia del problema, es posible decir que los usuarios de la BDCV conocen y emplean de manera adecuada el campo de búsqueda por temas.

4.1.1.4. Estado de la formulación

El problema más grave en las búsquedas temáticas se da en el plano semántico de la petición. Esto significa que las consultas de los usuarios no están debidamente

formuladas como para ser objeto de una búsqueda temática. Se contabilizaron 138 peticiones que presentan polisemia o cuyo significado no está correctamente delimitado.

Distintos artículos en donde se habla sobre el lenguaje natural señalan que éste se caracteriza por ser demasiado polisémico, en el sentido de que un símbolo bien puede referir a distintos conceptos y aún contextos. Ante este problema del lenguaje natural los lenguajes controlados, en su búsqueda por hacer más efectiva la recuperación de información, intentan eliminar la polisemia mediante la asociación de símbolos precisos para cada referente. Pese a lo anterior, este precepto de los lenguajes controlados es desconocido para los usuarios quienes llevan la polisemia del lenguaje natural a sus peticiones de búsqueda. Para una comprensión más precisa de lo enunciado, se exponen algunos casos: “Virus”, es una petición moderadamente polisémica, ya que puede referir a un “Virus (biología)” o un “Virus (informática)”. En comparación, otras peticiones pueden ser aún más polisémicas: “Día”, por ejemplo, es una consulta extremadamente ambigua, pues los encabezamientos brindan opciones como “Día de acción de gracias”, “Día de la raza”, “Día de las madres”, “Día de muertos”. Otro ejemplo es “Trastornos”, partimos de que no existe un encabezamiento con la palabra, sino más bien se usa compuesta con otras, así, forma encabezamientos como: “Trastornos cognitivos”, “Trastornos de adaptación”, “Trastornos de la memoria”, “Trastornos de la audición”, “Trastornos de la voz”, etc.

Ante estas consultas el sistema opta por un enlace de palabras clave, esto es, presenta al usuario todos los documentos que en alguna parte del encabezamiento tengan la palabra solicitada. Lo que redundará en una recuperación excesiva para el usuario, quien tiene que revisar muchos registros en busca de los que le son útiles. En el inconveniente detectado se hace evidente la falta de habilidad del usuario en la formulación de búsquedas por temas. Si asociamos estos resultados con otros estudios de recuperación de información, las cifras serían un indicador de que el catálogo de la biblioteca tiene un alto grado de “llamado” pero poca “precisión” en los documentos presentados. El problema, además, resulta alarmante por su alta recurrencia, obsérvese el porcentaje que registra:

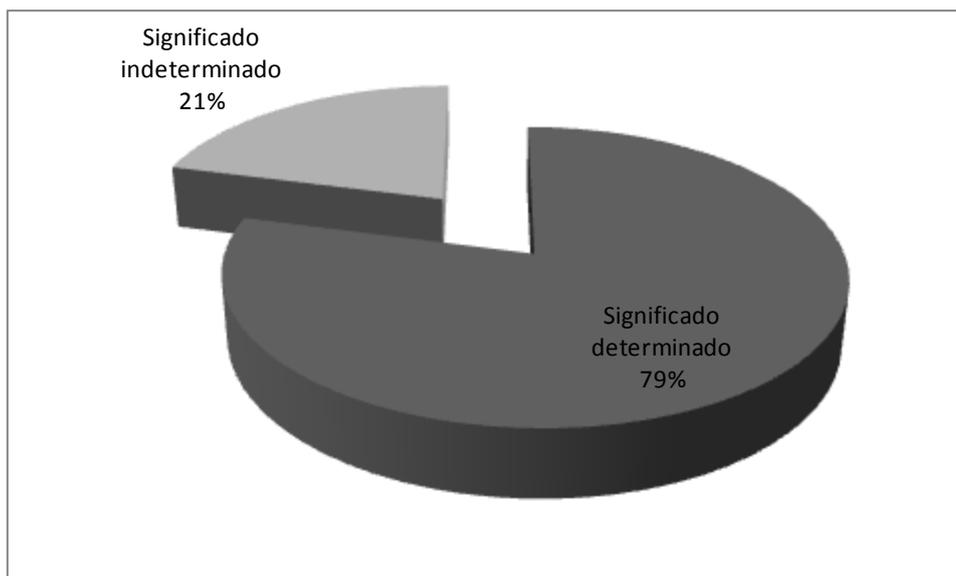


Fig. 12. Determinación del significado de la petición de búsqueda

De las 660 peticiones de búsqueda que integran la muestra el 21% no tienen un significado determinado del que pueda partir el catálogo para hacer un enlace preciso con el lenguaje de indización. Mientras que el lenguaje de indización busca ser lo más claro posible, las palabras empleadas por los usuarios en la recuperación carecen de ese rasgo. Además, mientras que los lenguajes controlados tienden hacia la estructura nominal como base, puesto que este tipo de palabras tienen referentes, los usuarios en sus búsquedas usan adjetivos, verbos y palabras cargadas de polisemia.

En ocasiones, las peticiones son, incluso, incomprensibles. Es en ese momento cuando surge la pregunta: ¿Qué es lo que el usuario pretendía buscar? Se incorporan algunos ejemplos que ayudan a entender lo dicho: “Límite”, “Prohibida” “Elaboración” “Monitoreo”. Un bibliotecario, ante consultas como estas, bien podría preguntar: ¿El límite de qué?, ¿Qué prohibida? ¿La elaboración de qué? o ¿El monitoreo de qué? Pero como la consulta va dirigida a un catálogo que no tiene la capacidad para discernir sobre si es adecuada o no, la máquina busca enlazar de algún modo con las palabras de indización.

La recuperación en las peticiones ambiguas es completamente imprecisa, por ejemplo, a partir de la palabra “Monitoreo” el sistema recuperó documentos con encabezamientos como “Monitoreo ambiental” “Ejecutivos-Monitoreo” y “Elecciones locales-Monitoreo”. ¿Alguna de estas opciones sería lo que el usuario buscaba? ¿Le habrán sido útiles los documentos recuperados? Saberlo no es posible, en virtud de que el análisis de transacciones no permite llegar hasta estos resultados. No obstante, casos

como estos abren la puerta a posteriores investigaciones que vinculen la precisión de la formulación de búsqueda y la utilidad de los documentos recuperados.

Las peticiones de búsqueda con significado ambiguo evidencian poca habilidad de los usuarios en la formulación de sus búsquedas temáticas. Ya en los libros especializados se ha señalado que la formulación de búsqueda es uno de los aspectos críticos de la búsqueda temática y este estudio lo reafirma. A la vez que apunta hacia la necesidad de ayudar a los usuarios para construir formulaciones de búsqueda temática, ya sea mediante cursos de formación en donde se les instruya sobre el proceso o bien se ligue la búsqueda por interrogación (Query) con la búsqueda por navegación (Browsing) mediante una vinculación con los índices en los cuales el usuario más que formular la búsqueda navega hasta encontrar el término más preciso a su necesidad de información.

4.1.2. Estructura

Uno de los objetivos específicos de la tesis busca saber cómo son las palabras que el usuario usa en su búsqueda por temas. Es un tema poco tratado, incluso en las norteamericanas que tratan más ampliamente la búsqueda temática, empero, resulta necesario saber qué tipo de palabras emplean los usuarios para buscar, y más aún, saber si hay relación entre la estructura de la petición y la estructura de indización.

La escasa literatura que puede brindar luces a la cuestión planteada menciona, aún con ciertas reservas, que la búsqueda temática empieza a tomar una estructura muy diferente a la que presentan los encabezamientos de materia. Uno de los investigadores que con más énfasis lo afirma es Ray Larson quien expone que: “en las solicitudes de búsqueda los usuarios no usan estructuras acordes con los encabezamientos de materias sino que sus búsquedas suelen ser parecidas a la búsqueda por palabras clave.”¹²² Otra investigación en torno al problema, advierte que “ha habido [...] un significativo y consistente declive en las búsquedas basadas en vocabularios controlados al paso de los años a favor de una búsqueda por palabra clave.”¹²³

Luego de analizar las peticiones de búsqueda en la BDCV, focalizando su estructura, se obtuvieron los siguientes resultados:

¹²² Ray Larson. “The decline of...*Op.cit.*”

¹²³ M. S. Sridhar. “Subject searching in the OPAC of a special library: problems and issues”. *OCLC Systems and Services: International Digital Library Perspectives* (Westport,CT.: Meckler, vol, 20, núm.4, 2004), pp. 183-191.

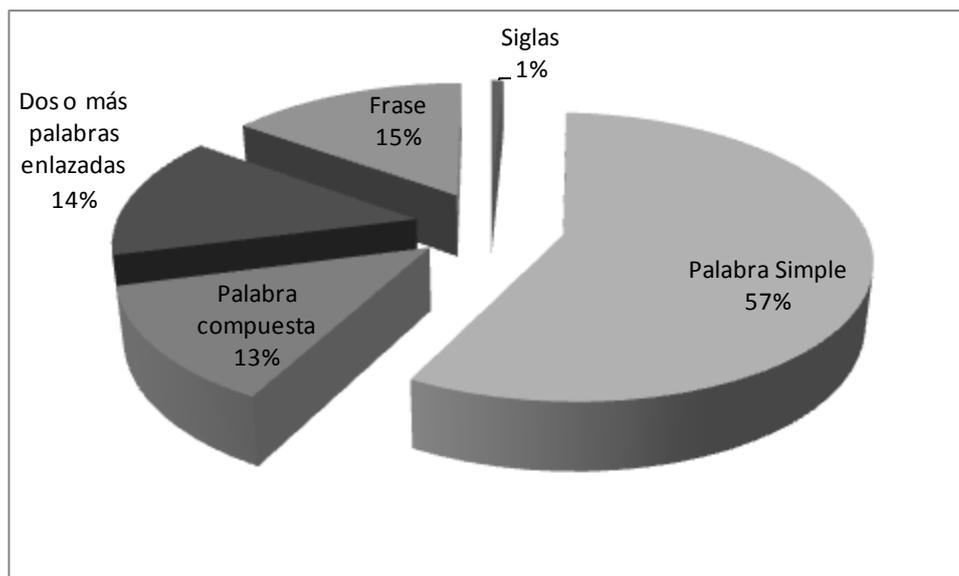


Fig. 13. Estructura de la petición de búsqueda

Como la teoría sugiere, la estructura más usada es la palabra simple. Le sigue la petición que usa una frase para representar el tema. Luego, dos o más palabras enlazadas. La estructura que ocupa el cuarto sitio es la formada por palabras compuestas y, finalmente, las siglas.

Los resultados coinciden con el trabajo de Sridhar, llevado a cabo en la biblioteca del ISRO Satellite Center, donde “la mayoría de los usuarios busca con un término único que rechaza el propósito de búsquedas combinadas o la lógica booleana.”¹²⁴ Del mismo modo, se observa que pocos usuarios se aventuran a buscar con más de una palabra, lo que implica el desuso de una estructura de búsqueda facetada.

4.1.2.1. Palabras simples

Son el 57% de la muestra. Su alta presencia afirma las especulaciones respecto a que la búsqueda temática ha tomado una estructura de palabras clave. Las peticiones expresadas en palabras simples pueden dividirse en dos grupos. El primero se forma de palabras simples que son imprecisas en la recuperación; y el segundo, se constituye de palabras simples que son precisas en ésta.

El primer grupo, palabras simples imprecisas, es más numeroso que el segundo; registra el 60% de los casos. Se integra por temas generales que el usuario bien podría delimitar para elevar la precisión en la recuperación. Ejemplos del grupo son: “Educación”, “Literatura”, “Historia”, “Economía”, “Filosofía” y “Religión”. Son temas

¹²⁴ *Idem.*

amplios y aceptados como encabezamientos que extraen demasiados documentos: “Educación”(13,436), “Literatura”(15,072), “Historia”(75,551), “Economía”(8,590), “Filosofía” (11,492) y “Religión” (3,822). Ante cifras tan elevadas, resulta poco probable que el usuario pueda revisar todos los materiales. Bien podrían señalarse como búsquedas no exitosas¹²⁵ por la sobrecarga de ítems obtenidos.

En distintos trabajos se ha dicho que el gran problema en las búsquedas temáticas es la falta de recuperación de materiales a partir de la petición de búsqueda. No obstante, otro de los problemas es la excesiva recuperación de documentos a partir de ésta. Sandy Roe¹²⁶, en una investigación llevada a cabo a principios de los noventa, señala ambos casos como dificultades a las que se enfrentan los usuarios. Mediante la asociación de estos casos a variables, Roe concluye que cuando el usuario emplea palabras muy generales la recuperación se incrementa y el usuario bien puede perderse en un mar de registros, en contraste, cuando se hacen búsquedas muy específicas se corre el riesgo de no recuperar documentos. Su recomendación es que el usuario emplee un término general al iniciar la búsqueda y lo delimite durante la interacción con el sistema.

El segundo grupo de palabras simples, formado por el 40% de las peticiones en palabras simples corresponde a consultas más delimitadas que recuperan un número adecuado de documentos. Ejemplos del grupo son “Pedofilia”, “Tanatología”, “Metacognición”, “Geopolítica”, “Alfabetización” y “Psicolingüística”. Se encontró que aproximadamente el 50% de estos términos son tecnicismos aceptados como encabezamientos en los lenguajes controlados, y aun cuando pueden tener facetas, éstas son menos en comparación con los temas del primer grupo.

4.1.2.2. Frases

El segundo tipo de petición más frecuente es la frase¹²⁷ con el 15% de apariciones. Mario Barite, sobre las frases, como estructuras para representar conceptos, señala que:

En los tesauros y en otros lenguajes documentales que utilizan conceptos para la representación del contenido se prefiere excluir la fraseología, aunque suele

¹²⁵ Karen Antell y Jie Huang. “Subject searching success... *Op.cit.* Definieron una búsqueda exitosa en términos de recuperación de documentos e indican que, desde el punto de vista de la recuperación, debe ubicarse entre 1 y 500 registros recuperados. Tomando como base ese estudio, se hizo la separación de palabras precisas con buena recuperación (exitosa) y palabras imprecisas con problemas de recuperación (no exitosa).

¹²⁶ Sandy Roe. “Online Subject Access” en *Journal of Internet Cataloging* (Binghamton, NY. : Hworth, vol.2, núm.1, 1999), pp. 69-78.

¹²⁷ Las frases son varias palabras enlazadas mediante conectores que no llegan a constituir una oración.

considerarse un porcentaje variable, pero relativamente alto, de sintagmas para envasar descriptores y no descriptores [...] Los teóricos de la clasificación recomiendan restringir al máximo el uso frases o sintagmas.¹²⁸

Los lenguajes controlados integran diversas estructuras: palabras simples, palabras compuestas, palabras enlazadas, frases, etc. Pero, la recomendación por excelencia en su construcción es eliminar en el mayor grado posible la fraseología con la finalidad de reducir la expresión temática a estructuras sencillas que eviten la sinonimia y la polisemia.

En el corpus analizado las búsquedas expresadas en frases son muy frecuentes. Esto hace evidente una separación entre los principios de los lenguajes de indización y el comportamiento de los usuarios en la búsqueda por temas. Dado el porcentaje de casos (15%) deben tomarse medidas. Se vislumbra como solución un estudio en el que se trate el comportamiento del usuario en la búsqueda por temas y, después, se diseñe un proyecto de formación dirigido a campos de búsqueda específicos.

En la investigación se encontró que las consultas mediante frases son sintagmas largos, ejemplos de ellos serían: “Archivo histórico del agua”, “Características nacionales cubanas”, “Internet en la educación superior” “Estudios sobre la familia mexicana”, “Postulados generales de la ciencia” e “Historia del evolucionismo”.

4.1.2.3. Dos o más palabras enlazadas

La petición formada por dos o más palabras enlazadas ocupa el tercer sitio con el 14% de apariciones. Son estructuras en las que el usuario liga dos ideas conceptualmente distintas de las que quiere obtener información. Ejemplos del grupo son: “Diccionarios and derecho”, “Alcazar and Segovia”, “Frases célebres and historia de México”. Otras veces no está presente el operador booleano; no obstante, se aprecian las ideas que se intenta ligar: “Bolivia siglo xix”, “Comunicación estereotipos”, “Muerte edad media” y “Derecho aspectos políticos”. En las consultas se percibe que los usuarios intentan delimitar el tema de algún modo y para hacerlo añaden otra palabra. Su recuperación muy buena, especialmente cuando se usan operadores booleanos que restringen o amplían la búsqueda.

¹²⁸ Mario Guido Barite Roqueta. *Los conceptos y su representación...Op.cit.* p. 35.

4.1.2.4. Palabras compuestas

El cuarto sitio con el 13% de las peticiones está formado por palabras compuestas. Son términos que refieren a un concepto delimitado por un adjetivo. Se muestran algunos ejemplos: “Educación básica”, “Revolución mexicana”, “Opinión pública”, “América latina”, “Bacterias antagonistas”, “Libro electrónico”, “Números primos” y “Sistema educativo mexicano”. Las consultas eliminan la ambigüedad semántica, es decir, no piden que el sistema busque “Educación”, sino una mejor opción: “Educación básica”. Esta delimitación restringe el número de documentos detectados por el sistema y por tanto la recuperación se torna precisa.

4.1.2.5. Siglas

Son las peticiones menos frecuentes, con una presencia de 1% en el corpus. Es común que muchas instituciones, tratados u otro referentes puedan ser nombrados por siglas y sean, en ocasiones, más conocidos por éstas que por su nombre completo. Sobre ellas, la teoría de indización establece que debe registrarse como tema el nombre completo de la entidad y las otras formas, entre ellas siglas, deben recogerse como temas no autorizados bajo un registro de autoridad.

En la muestra se registraron sólo cinco casos: “IFE”, “BANAMEX”, “IETU”, “TLCAN” y “UNIFEM”. De los casos, dos recuperan documentos “IFE” y “BANAMEX”, mientras que los otros tres: “IETU”, “TLCAN” y “UNIFEM” no obtuvieron materiales, aun cuando existe un registro de autoridad para la forma completa. La BDCV tiene una política donde señala que las formas cortas de un nombre deben ser incorporadas en los registros de autoridad, por lo tanto, estas omisiones son errores del área de catalogación por temas.

4.1.2.6. Impacto de la estructura en la recuperación de documentos

Una vez obtenida la estructura de la petición de búsqueda se pretendió saber si había alguna relación entre la estructura de las peticiones de búsqueda y la recuperación de documentos. Se consiguieron datos interesantes.

En las búsquedas que recuperaron documentos la estructura quedó como sigue:

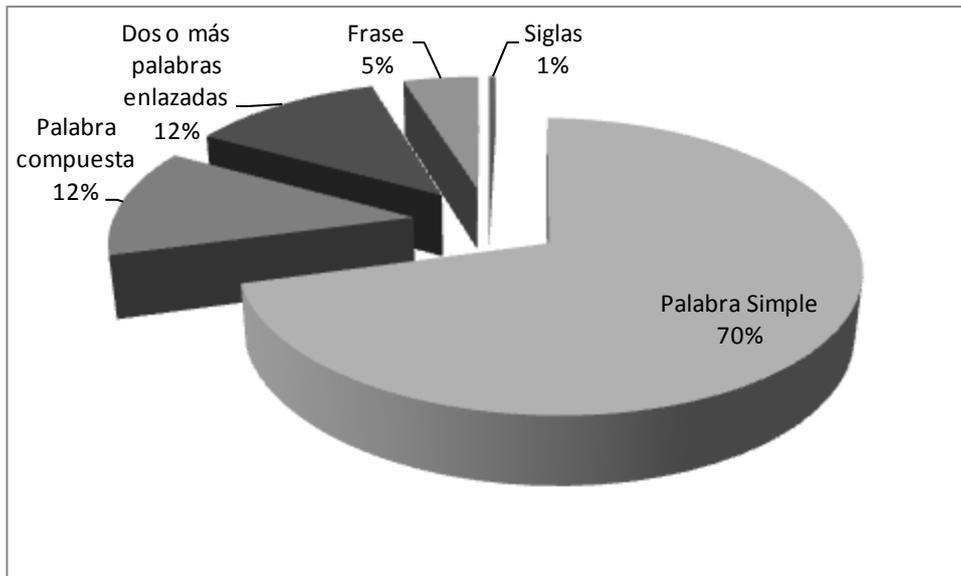


Fig. 14. Estructura de la petición en búsquedas que recuperan documentos

El 70% son búsquedas a partir de una palabra. Con el mismo porcentaje le siguen las peticiones de palabras compuestas y las de dos o más palabras enlazadas. Las frases que recuperan documentos son el 5% y las siglas son el 1%.

Se presentan ahora, los resultados de las búsquedas que no obtuvieron documentos:

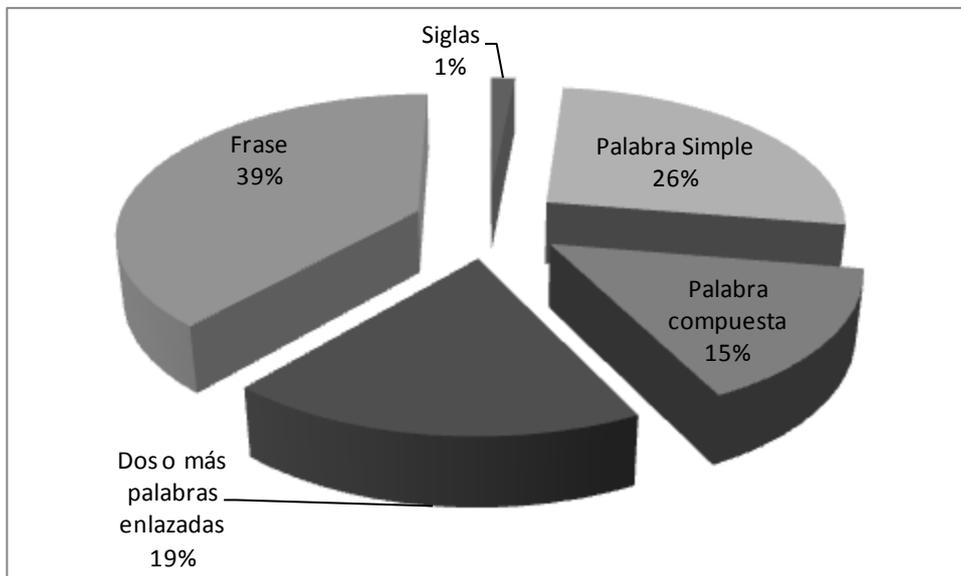


Fig. 15. Estructura de la petición en búsquedas que no recuperan documentos

Como se puede observar, la estructura dominante es la frase. Le siguen las palabras simples, dos o más palabras enlazadas, palabras compuestas y siglas. Se concluye que la búsqueda donde no hay recuperación está vinculada al uso de frases en el momento de la formulación de búsquedas, y sobre estas, es posible afirmar que no

son términos de búsqueda sino meras descripciones de la necesidad del usuario. Se exponen algunos ejemplos que sirven para la comprender lo enunciado: “Clasificación de empresas según la forma de explotación”, “El imperio romano antes de la era cristiana” e “Impacto del comercio intraindustrial en la industria manufacturera”. En la solicitud se comprueba que el usuario tiene una idea muy clara de lo que necesita, incluso, si le planteara esta necesidad a un bibliotecario de consulta o referencia obtendría mejores resultados que buscando en el catálogo. El motivo de la búsqueda fallida radica en que, frente al catálogo, el usuario debe llevar su necesidad hasta la formulación de símbolos comprensibles para el sistema. Cito a Chowdhury quien dice que: “las necesidades del usuario necesitan ser trasladadas y puestas en el sistema de recuperación en el mismo lenguaje de éste usando los mismos términos, por ejemplo que fueron usados para expresar el contenido de los documentos registrados.”¹²⁹ Las frases introducidas en la solicitud se encuentran en un estado de necesidad consciente, no atraviesan el umbral de la formulación y menos aún representan símbolos comprensibles para el SRI.

4.1.3. Características

Las peticiones temáticas pueden ser de tres tipos: tópicas, personales y geográficas. La proporción de estos tipos en el corpus fue el siguiente:

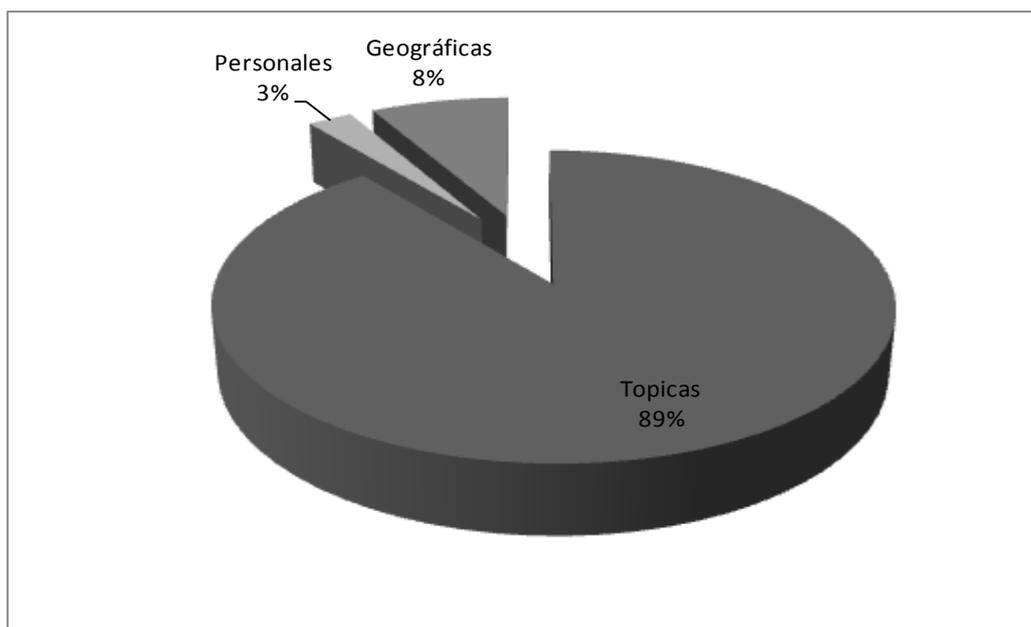


Fig. 16. Tipo de petición

¹²⁹ G.G. Chowdhury. *Introduction to modern information retrieval* (Londres: Facet Publishing, 2004), p. 123.

Las más frecuentes son las de t3pico, le siguen las geogr3ficas y en 3ltimo lugar las personales.

4.1.3.1. B3squedas por t3pico

Las peticiones por t3pico son la mayor3a del conjunto, corresponden al 89% de las solicitudes de b3squeda. Se percibe que los usuarios asocian la b3squeda por tema con su naturaleza t3pica, a3n cuando la biblioteca hace entradas tem3ticas para nombres personales, nombres corporativos y entradas tem3ticas geogr3ficas.

De las peticiones t3picas se hizo una divisi3n en cuatro categor3as: peticiones bien escritas que coinciden con un encabezamiento de materia; peticiones bien escritas que no coinciden con un encabezamiento de materia; peticiones mal escritas que coinciden con un encabezamiento de materia y peticiones mal escritas que no coinciden con un encabezamiento de materia.

Seg3n la clasificaci3n establecida, se obtuvieron los siguientes resultados:

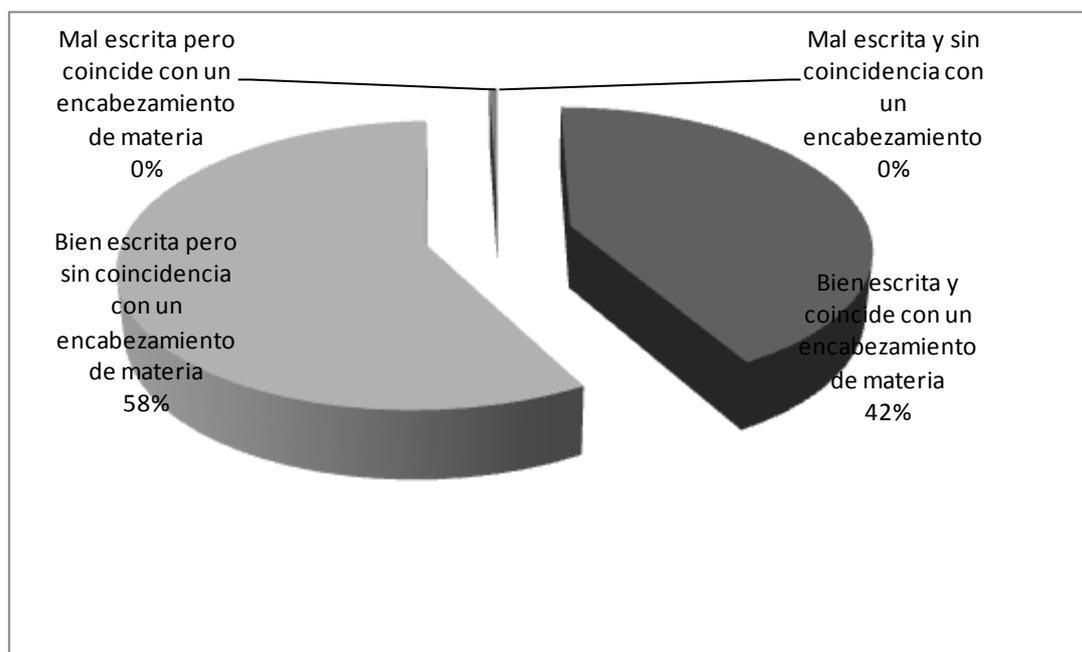


Fig. 17. Peticiones t3picas

El 42% de las peticiones t3picas est3n bien escritas y coinciden con un encabezamiento de materia asignado por la BDCV. Casos representativos son: “Cristianos”, “Musulmanes”, “Masoquismo”, “Olimpiadas”, “Feminismo”, “Fenomenolog3a”. Gran parte de ellas, 123 de 169, es decir el 73% son palabras simples. As3, el lenguaje del usuario tiene mayor grado de coincidencia con el lenguaje de indizaci3n cuando 3ste formula su petici3n de modo general.

En menor proporción, en el 27% se presentan peticiones con estructuras más complejas que coinciden con encabezamientos temáticos. Se exponen algunos casos: “Periodismo and México and Historia”, “Internet en la educación superior”, “Tecnologías de la información”, “Salud pública and Congresos”, “Comunicación and Aspecto sociales and Congresos”, “Seguridad ciudadana”, “Revoluciones and América Latina”. Las solicitudes más complejas que coinciden con encabezamientos se caracterizan por tener una estructura más cercana a la del lenguaje de indización, el uso de operadores booleanos y el uso de facetas de búsqueda. Lo que denota mayor habilidad de ciertos usuarios en la construcción de la petición.

Las peticiones tópicas bien escritas pero sin coincidencia con un encabezamiento de materia son el 58% del total. Es una cifra mayor que la coincidencia. De ello se concluye que las peticiones tópicas en su mayoría no coinciden con el lenguaje de indización.

En el tercer rubro de las peticiones tópicas se ubican aquellas que están mal escritas pero coinciden con un encabezamiento de materia. Sólo se detectaron dos casos: “Hospicios” y “Meixoco” que corresponden a los encabezamientos “Hospicios” y “México”. La literatura señala que muchas búsquedas fallidas son causadas por problemas ortográficos y suponen que colocar un corrector ortográfico en el motor de búsqueda solucionaría el problema. En el caso de la BDCV, efectivamente, solucionaría el problema, pero en función de la baja presencia del inconveniente en el acceso por materias no resulta necesario.

El cuarto grupo, peticiones mal escritas y que no coinciden con un encabezamiento de materia no tuvo presencia en la muestra analizada.

4.1.3.2. Búsquedas personales

Las entradas personales como peticiones temáticas tienen poca recurrencia, son apenas el 3% del corpus. Las categorías bajo las que se agrupan sólo son dos de las cuatro establecidas para el análisis: bien escritas que coinciden con un encabezamiento y bien escritas que no coinciden con un encabezamiento:

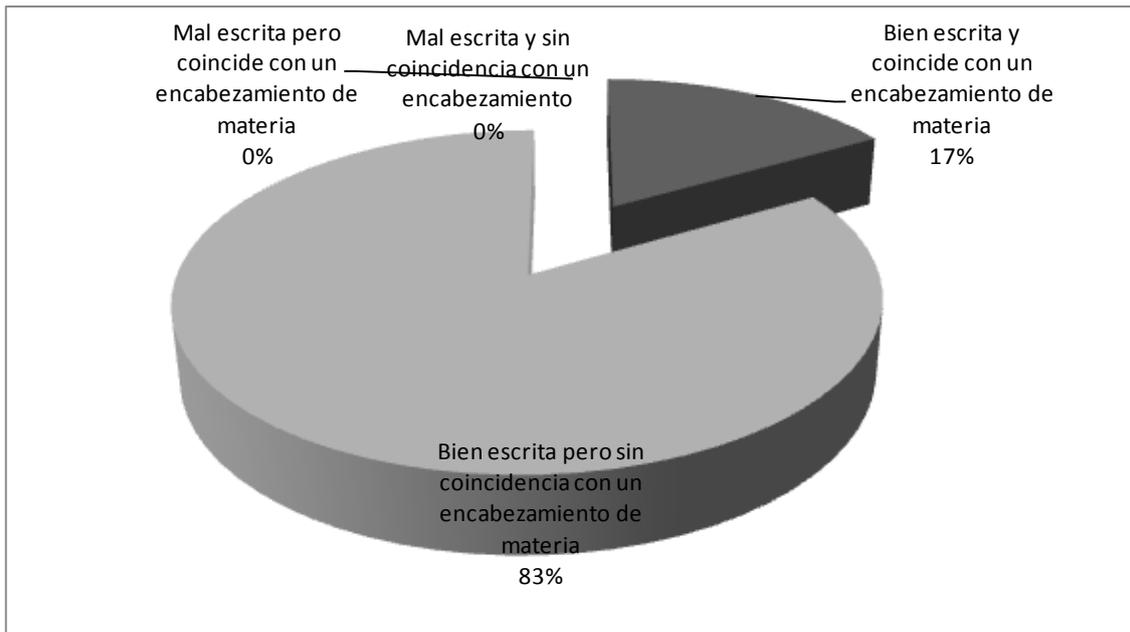


Fig. 18. Peticiones personales

Las correctamente escritas y con coincidencia son dos: “González, Felipe” y “González y González, Luis”. Mientras que las escritas correctamente, pero que no coinciden con un encabezamiento, como debe ser asentado, son 10 casos. La no coincidencia en las peticiones temáticas bajo persona más bien es una alteración en la estructura de los encabezamientos personales que no impide la recuperación. Corresponden a búsquedas donde los usuarios anteponen el nombre del autor al apellido: “Hermenegildo Bustos”, “Daniel Cosío Villegas” “Andrés Quintana Roo”. En los catálogos en tarjetas el hecho implicaba que el usuario no recuperara documentos, pero con la llegada de los catálogos electrónicos el inconveniente se ha visto superado.

Otra característica de las consultas personales es la imprecisión en la petición. En casos como “Obregón”, “Santa Anna” o “Kant” el problema es superado, pues a partir de los datos almacenados el sistema establece principios de relevancia y extrae gran cantidad de ítems sobre: “Obregón, Álvaro, pres. México, 1880-1928”, “Santa Anna, Antonio López de, pres. México, 1794-1876” y “Kant, Immanuel 1724-1804”. En contraste, hay otras peticiones ambiguas donde en las que el sistema no pudo establecer principios de relevancia. Es el caso de “William”, “Fernando”, “Felipe” “Serrano”, “Valle” “Garrido” o “Truman”. Es evidente que el usuario intentó buscar por una entrada temática bajo nombre personal, pero ante el desconocimiento del nombre completo la búsqueda arrojó muchos resultados sobre personajes cuyos nombres incluía alguna o algunas de las palabras solicitadas.

4.1.3.3. Búsquedas geográficas

Su presencia es un poco mayor que las búsquedas temáticas bajo entrada personal; alcanza una presencia del 8% de la muestra. Se agrupan, igual que los encabezamientos personales, bajo dos categorías: las que están bien escritas y coinciden con un encabezamiento geográfico y las bien escritas que no tienen estructura de encabezamiento:

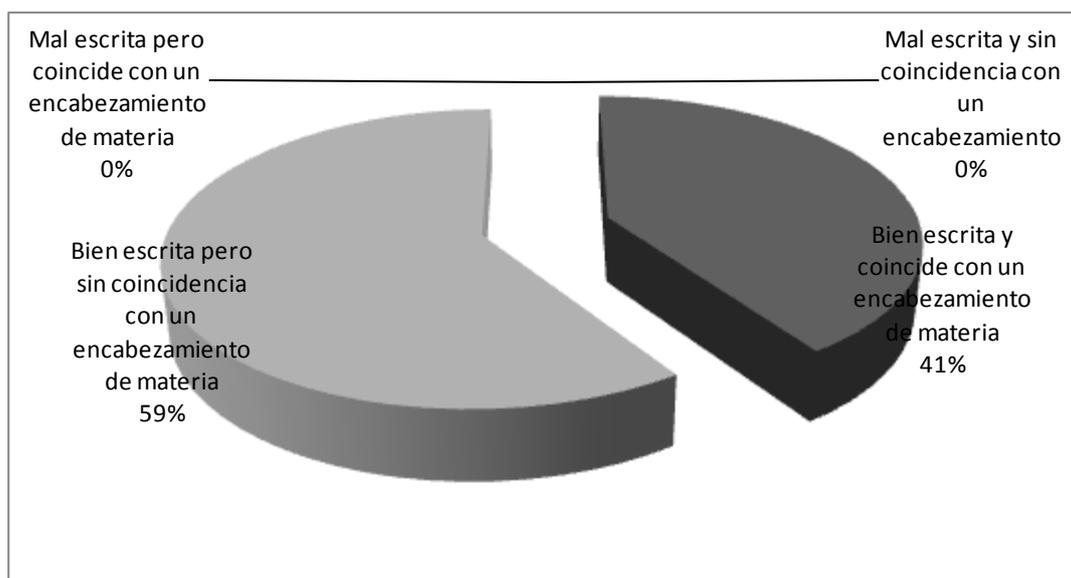


Fig. 19. Peticiones geográficas

Entre las primeras se encuentran solicitudes como “Brasil”, “América”, “Asia”, “África”, “Vietnam” y “Tlaxcala”. En las segundas, bien escritas pero que no coinciden con un encabezamiento se insertan aquellas en las que el usuario busca información de una región geográfica más específica o en las que agrega otra faceta de búsqueda. Por ejemplo: “Revolución México”, “Haciendas Puebla”, “Historia Argentina” y “Cholula urbanismo ciudades colonia”. Son casos en los que la entrada principal debió haber sido geográfica y, luego, debieron ser delimitados por subencabezamientos para accesos geográficos. Así, para ser objeto de coincidencia con el lenguaje de indización las peticiones debieron formularse como: “México-Revolución, 1910-1920”, “Puebla-Haciendas”, “Argentina-Historia” y “Cholula, Puebla-Antiguedades”. Nuevamente, el problema radica en que son búsquedas expresadas bajo principios de lenguaje natural.

4.2. Coincidencia de la petición de búsqueda con la indización

Las peticiones de los usuarios obtienen distintas respuestas del sistema, en términos de recuperación de documentos. Muchos artículos especulan sobre las cifras precisas de enlace o coincidencia entre las peticiones del usuario, expresadas en lenguaje natural, y la indización, llevada a cabo en lenguajes controlados.

En la presente investigación los resultados fueron los siguientes:

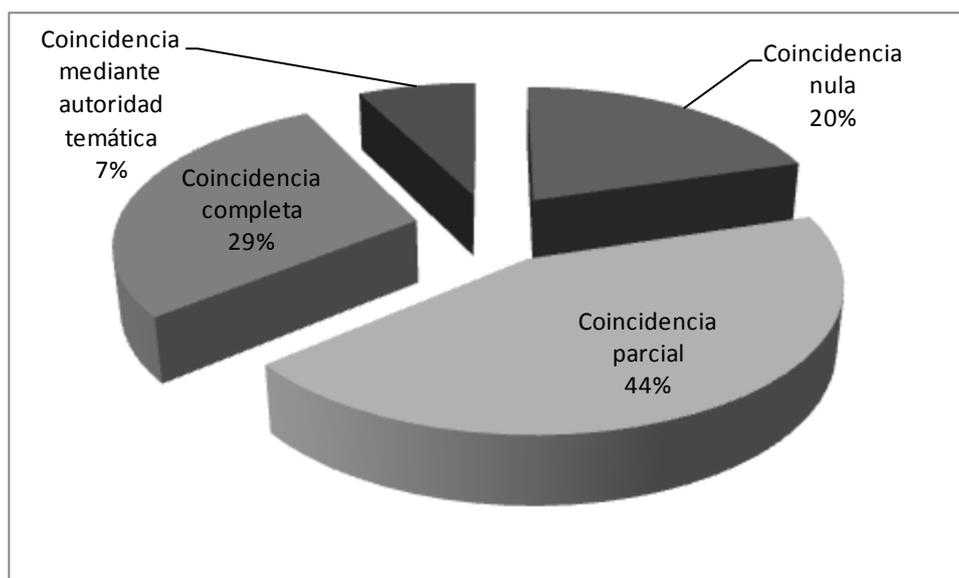


Fig.20. Coincidencia entre la petición y la indización

Como se observa, la más frecuente es la coincidencia parcial, le siguen la coincidencia completa, la coincidencia nula y, finalmente, la coincidencia mediante autoridad temática.

4. 2.1. Coincidencia nula

La falta de coincidencia entre las consultas del usuario y la indización es considerada el principal problema de la recuperación por temas. Al respecto, Rumi Y. Graham indica que: “solemos pensar que las búsquedas con cero resultados son insatisfactorias.”¹³⁰ Estos casos registran el 20% de la muestra. Son solicitudes que de ninguna manera pueden ser enlazadas con la indización. Ejemplos son: “Educación de prisioneros”, “Crimen y mafia italiana”, “Cuerpo y expresión”, “El turismo accesible”, “El imperio romano antes de la era cristiana”, “Empleadores de personas con discapacidad”,

¹³⁰ Rumi Y. Graham. “Subject No-Hits Searches... *Op.cit.*”

“Enseñanza universitaria-recursos en redes de comunicación” y “La fenomenología de la religión”. Son peticiones que tienen cobertura temática en la colección; se descarta, por tanto, que no haya documentos sobre el tema. Aunado a esto, son peticiones con semanticidad, esto es, su significado es comprensible. Luego de analizarlas, se determinó que la no recuperación se debe a tres causas: estado inmaduro de la formulación de búsqueda, limitaciones léxicas del lenguaje controlado frente al lenguaje natural y separación de la estructura del lenguaje natural con la estructura del lenguaje controlado.

Sobre el primer aspecto, formulación de búsqueda, una consulta pasa por tres etapas:¹³¹ un primer momento donde en la mente se tiene un referente, un segundo lapso en el que se tiene el concepto del referente y un último estado cuando el concepto se concreta en un símbolo. Muchas de las búsquedas fallidas están en el plano del concepto expresado en frases y nunca llegan a la etapa del símbolo expresado en términos comprensibles para el sistema. Es como si un usuario dijera que quiere un libro sobre “la economía de México en la década de los ochentas”, pero nunca llegase al símbolo: “México- Aspectos económicos- 1980-1990”. Son peticiones mantenidas en lenguaje natural que nunca fueron “traducidas” al lenguaje controlado y por eso, el catálogo no tiene respuesta a ellas. Se resume que así como un catalogador “traduce” el aboutness de un libro en términos de indización, el usuario debe “traducir” su necesidad en términos de recuperación.

La segunda causa de estas búsquedas fallidas se debe a limitaciones léxicas del lenguaje controlado frente a peticiones del usuario planteadas en lenguaje natural. El lenguaje controlado tiene una estructura muy organizada en la que, según la teoría sobre organización de la información, todo conocimiento tiene cabida. Sin embargo, resulta imposible que un lenguaje controlado albergue todas las posibilidades de expresión (variables léxicas) de un referente expresado en lengua natural. Recoger todas las manifestaciones del lenguaje natural es una tarea casi imposible. Aunado a esto, la lengua natural tiene tantos regionalismos que sería necesario crear lenguajes controlados acordes a cada región.

La tercera causa de la no recuperación de documentos es la separación de la estructura del lenguaje natural con la estructura del lenguaje controlado. Los usuarios desconocen la sintaxis propia de los lenguajes controlados, específicamente la

¹³¹ Según el triángulo de expresión de Charles Kay Ogden y I.A. Richards. *El significado del significado...Op.cit.* p.339-340.

estructura de los encabezamientos de materia. Se determinó que gran parte de las peticiones tienen una sintaxis característica de los buscadores de Internet donde es posible la coincidencia entre la solicitud y partes completas del documento. En un trabajo similar, sobre otro catálogo mexicano (LIBRUNAM) Juan Manuel Sánchez Zurita¹³² advierte que entre los usuarios hay una tendencia a buscar como en *Google*, es decir, “en posibilidades de búsqueda que el motor permite: la búsqueda libre por palabras (lenguaje natural)”.

Los estudios que han detectado el problema de la nula recuperación de documentos han sugerido incorporar las peticiones sin resultados a una autoridad mediante la cual puedan ligarse a términos autorizados. Los registros de autoridad serían útiles para resolver dificultades de las búsquedas por autor y título, donde es posible predecir las manifestaciones que puede tener un asiento; no obstante, en la búsqueda temática el panorama se abre como un abanico de posibilidades cuando nos planteamos las opciones con las que el usuario podría buscar un tema.

4.2.2. *Coincidencia parcial*

En este rubro se registró el 44% de las peticiones totales. La coincidencia parcial es un problema muy grave en la recuperación en virtud del modo como se da. Si un usuario introduce la petición “Cuerpo y comunicación” el sistema busca un enlace completo, pero, al no encontrarlo, enlaza de algún modo con el lenguaje de indización. De modo que divide la estructura y empieza a rastrear por palabras clave “Cuerpo” y “Comunicación” presentes en cualquier parte de un encabezamiento. Al hacer esto, el catálogo presenta documentos que tratan sobre los temas desligados. Así, el usuario recibe documentos cuyos encabezamientos son “Mente y cuerpo”, “Comunicación oral” y “Comunicación no verbal”. Se intuye que al menos el documento encabezado bajo “Comunicación no verbal” le servirá al usuario. Pero, hay otros ejemplos de coincidencia parcial donde los resultados se alejan conceptualmente de la petición, tal es el caso de “Estudios sobre la cultura”. El sistema intenta una coincidencia completa con la indización y, al no ser posible, parte la estructura y busca “Estudios” y “Cultura”. Luego de hacer esto, recupera un documento que tiene los encabezamientos “Mujeres en la cultura- América Latina” y “Estudios sobre la mujer- América Latina”. Lo que se

¹³² Juan Manuel Sánchez Zurita. *Los usuarios Google frente al catálogo bibliotecario de la UNAM* (México: UNAM, Dirección General de Bibliotecas, 2009. (DE, 30 de Septiembre, 2009: <http://juanzurita.wordpress.com>)).

dio fue una recuperación parcial que, al desfragmentar la solicitud, recupera documentos inútiles.

La configuración del buscador de la BDCV parece regirse por el principio de recuperación. Este hecho resulta loable si lo que queremos es darle documentos al usuario; no obstante, está dejando de lado el factor precisión y relevancia de los documentos recuperados.

4.2.3. *Coincidencia mediante autoridad temática*

Durante los últimos años, la BDCV ha llevado a cabo un proceso de creación y fortalecimiento de autoridades. La utilidad de esta herramienta ayuda para que el 7% de las peticiones expresadas en lenguaje natural sean dirigidas a un encabezamiento del vocabulario controlado. Ejemplos de ellas son: “Cárceles” redirigida a “Prisiones”, “Didáctica” a “Enseñanza”, “Gay” a “Homosexuales”, “Misterio y cuentos” a “Cuentos policíacos y de misterios”, “Prisioneros” a “Presos” y otros casos similares donde el catálogo de autoridades está siendo útil para establecer un enlace entre la solicitud de búsqueda y la indización.

4.2.4. *Coincidencia completa*

Los estudios que indagan sobre la coincidencia entre el lenguaje de indización y el lenguaje del usuario han obtenido cifras muy bajas respecto a la coincidencia entre ambos. Los resultados oscilan entre el 14% y el 45% de los casos. Esta investigación reafirma lo mencionado por dichos estudios. La coincidencia completa entre el lenguaje del usuario y el lenguaje de indización se presenta en el 29% de la muestra. Esto implica que menos de una tercera parte de las peticiones son objeto de una recuperación precisa respecto a la consulta del usuario.

La coincidencia completa se presenta en temas generales. Algunos de los casos que se registran son “Psicoanálisis”, “Prensa”, “Revolución”, “Periodismo”, “Paz”, “Matrimonio”, “Urbanismo”, “Oasis”, “Soberanía”. Marcia Bates¹³³ señala los beneficios de partir de consultas generales y delimitar el tema con ayudas proporcionadas por el catálogo, empero, en la BDCV esto no sucede. Aunque se cuenta con índices y un catálogo de autoridades, éstos no se encuentran enlazados a la búsqueda por temas. En este sentido, la biblioteca debería valorar la posibilidad

¹³³ Marcia Bates. “System meets user...”. *Op.cit.*

enlazarlos para posibilitar a los usuarios una búsqueda por navegación ligada a la búsqueda por petición.

4.3. Conclusión

Mediante este instrumento fue posible concluir que el encuentro entre el lenguaje de indización y el lenguaje del usuario presenta problemas en dos momentos: el momento de la solicitud de búsqueda y el momento de enlace entre la petición y la indización.

Los problemas de la petición de búsquedas se relacionan con:

- a) Idioma: la solicitud de búsqueda y la indización están en lenguas distintas.
- b) Naturaleza de la colección: la petición de búsqueda no registra cobertura temática dentro de la colección de la biblioteca.
- c) Estado de la formulación: la petición se encuentra en un estado inmaduro de formulación y por tanto el sistema no puede dar respuesta a ella. Es un problema de habilidades del usuario para formular búsquedas por temas.
- d) Estructura: son peticiones cuya estructura, expresada en lengua natural, es completamente distinta a la estructura de la indización, correspondiente a la de los lenguajes controlados. Este hecho causa que no haya posibilidad de empalme entre una y otra, lo que repercute en la recuperación.

Los problemas de coincidencia entre la solicitud de búsqueda y la indización son graves: un 20% tiene coincidencia nula, el 44% tiene coincidencia parcial caracterizada por la falta de precisión y relevancia de los documentos recuperados, el 29% coincide completamente y el 7% restante corresponde a la coincidencia enlazada a través de una autoridad temática. Los resultados coinciden con investigaciones similares donde se indica la falta de enlace entre las consultas del usuario y la organización llevada a cabo por la biblioteca. También se percibe una gran separación entre el modo como indiza la biblioteca y la manera como busca el usuario, directamente relacionado con el lenguaje empleado en las actividades y con la falta de formación de los usuarios en las búsquedas por temas. Para que la búsqueda por temas sea más efectiva será necesario difundir entre los usuarios los beneficios de este campo de búsqueda y formarlos en el uso de lenguajes documentales y el empleo del catálogo. Sólo en la medida que instruyamos al usuario éste será más autónoma y sacará mayor provecho de el trabajo de los catalogadores.

**Capítulo 5. Resultados obtenidos mediante el cuestionario
sobre la búsqueda por temas**

5.1. Uso de la búsqueda

Actualmente las bibliotecas se preguntan si la búsqueda por temas es usada o no. La bibliografía de principios de los ochenta y hasta principios de los noventa¹³⁴ señalaba que aproximadamente la mitad de las búsquedas en bibliotecas era mediante este campo. Sin embargo, investigaciones recientes apuntan sobre el desuso de la opción. Ante tales aseveraciones resulta necesario determinar su estado en la BDCV, pues en ella se invierten recursos que tal vez no estén siendo útiles.

En la biblioteca se encontró que la opción de búsqueda más usada es la básica. El 84% de los usuarios la emplea, mientras que el 16% restante opta por emplear búsquedas avanzadas.

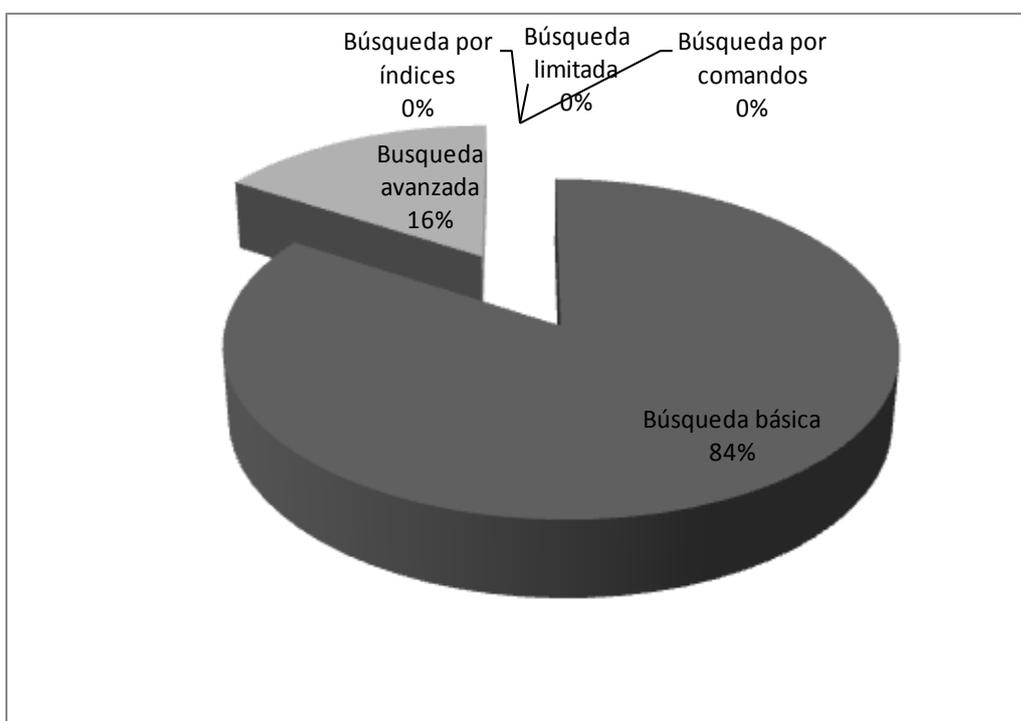


Fig. 21. Uso de tipos de búsqueda

En lo referente al uso de los distintos campos de búsqueda que ofrece el catálogo, este trabajo es de los primeros en ofrecer luces entorno al cuestionamiento. Se obtuvieron los siguientes resultados:

¹³⁴ *Idem.*

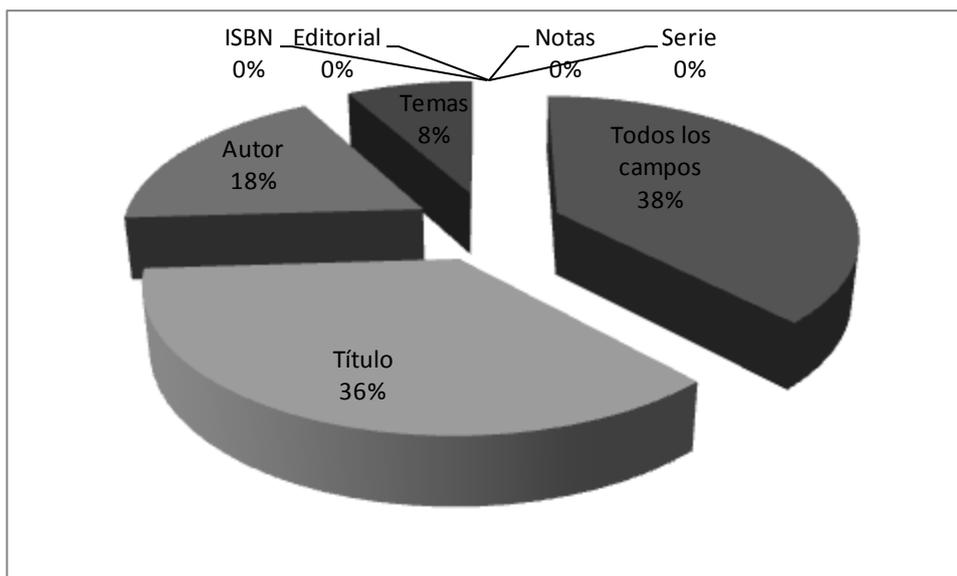


Fig. 22. Uso de los campos de búsqueda

Los campos más usados son: título y todos los campos. La búsqueda por temas, como muchas investigaciones recientes señalan, se ubica en el cuarto sitio, detrás de las opciones todos los campos, título y autor. Llama la atención el posicionamiento de la opción todos los campos que iguala a la opción de búsqueda por título y supera las opciones de autor y tema. Estos datos nos indican que los usuarios de la BDCV, por un lado, están interesados en recuperar información precisa a partir de un título con el que de antemano cuentan o bien y, por otro, buscan por todos los campos, porque como ellos mismo señalan es una opción “fácil”, “arroja muchos resultados”, “al lanzar la búsqueda en todos los campos se obtienen buenos resultados” y “porque muchas veces no tienen el autor o título del libro.”

La búsqueda por temas ocupa el cuarto sitio, con un 8%. De los puntos de acceso principales (título, autor y tema) es el menos usado. Este dato parece alarmante si lo vemos de manera aislada; sin embargo, otras investigaciones que se centran en la búsqueda temática han obtenido resultados similares. Por ejemplo, Karen Antell¹³⁵ quien hizo su trabajo en la biblioteca de la Universidad de Oklahoma menciona que de 14,234 transacciones usadas para su trabajo sólo 650, es decir, el 4.6% eran búsquedas temáticas. Además, eran el porcentaje más bajo detrás de las búsquedas por autor, palabra clave, título de revista y título. Los resultados de distintas investigaciones, en efecto, nos indican que la búsqueda temática está siendo poco usada y va en declive frente a otras opciones de búsqueda.

¹³⁵ Karen Antell y Jie Huang. “Subject searching success...”, *Op.cit.*

La presente investigación partió de que la búsqueda temática era una de las opciones más empleadas; sin embargo esto no es así. Ante el porcentaje de uso tan bajo que se obtuvo, se consultó la bitácora de búsquedas de Aleph durante tres semanas. Se hallaron estos resultados: la búsqueda en todos los campos obtuvo los siguientes porcentajes de uso: 37.68%, 43.77% y 38.27%. La búsqueda por título registro 28%, 26.58% y 25.47%. La búsqueda por autor 10.28%, 10.37% y 10.26%. En cuarto lugar se ubicó la búsqueda por temas con 3.18%, 2.68% y 4.50%. El resto de las opciones de búsqueda: Serie, ISBN, Notas, Editorial, Año, ISSN, Idioma, Catalogador, Fecha de captura, Fecha de catalogación, Fuente de catalogación, Número de control, Número de sistema, Texto completo, Lugar de publicación y Clasificación, registran porcentajes de uso muy bajos.¹³⁶

Los resultados de la bitácora de Aleph coinciden con los del cuestionario y llevan a concluir que la búsqueda por temas es el cuarto campo empleado, por detrás de todos los campos, título y autor y por encima de otros campos usados esporádicamente.

Al interrogar a los usuarios, directamente, acerca de si hacen uso de la búsqueda por temas, se obtuvieron estos resultados:

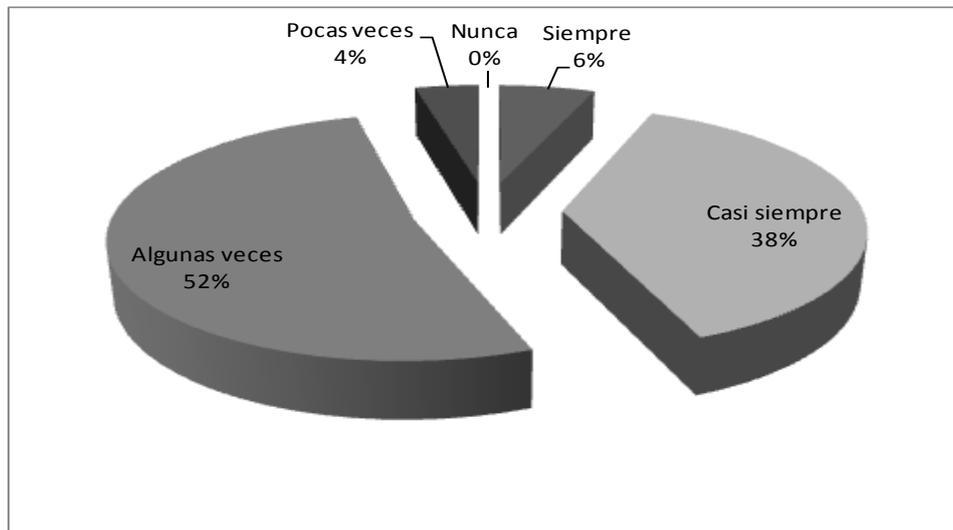


Fig. 23. Uso de la búsqueda por temas

La respuesta más generalizada es que la usan algunas veces. Sobre la frecuencia de su uso se encontró que quienes la usan dicen utilizarla en una etapa de búsqueda bibliográfica, generalmente, al principio de una tarea o investigación, cuando están buscando bibliografía. Los que respondieron usarla casi siempre lo hacen porque arroja

¹³⁶ En promedio, estos campos no superan el 1% de las búsquedas generales.

gran cantidad de resultados de los que pueden ir seleccionando los más adecuados a su necesidad de información. Quienes la usan algunas veces mencionan dos factores asociados con el uso: cuando no tienen la referencia de un documento preciso o cuando requieren más documentos para su trabajo y no tienen las referencias exactas. Los que mencionaron usarla pocas veces argumentan que es una búsqueda poco precisa en la que deben ir revisando documento por documento y, finalmente, quienes dicen no usarla argumentan que otras opciones de búsqueda (título, autor y todos los campos) ofrecen mejores resultados.

Con base en las respuestas abiertas donde los usuarios argumentaron por qué usan o no usan la búsqueda por temas, se concluye que los usuarios de la BDCV prefieren buscar por título a autor cuando cuentan con el dato exacto, pero si no tienen el dato exacto optan por hacer un rastreo en todos los campos. Sólo cuando las opciones de búsqueda anteriores han sido descartadas y no han recuperado documentos, deciden buscar por temas.

5.2 Opinión de los usuarios sobre la búsqueda

Las investigaciones que alertan sobre el bajo uso de la búsqueda por temas señalan como factor crítico del desuso la dificultad que implicar realizarla, pues dicen que en la búsqueda por título y autor el usuario ya cuenta con el dato a buscar mientras que en la búsqueda por temas implica un proceso de formulación de la consulta. No obstante, en la BDCV esto no es así, para los usuarios la búsqueda por temas no es difícil. Al respecto se obtuvieron los siguientes resultados:

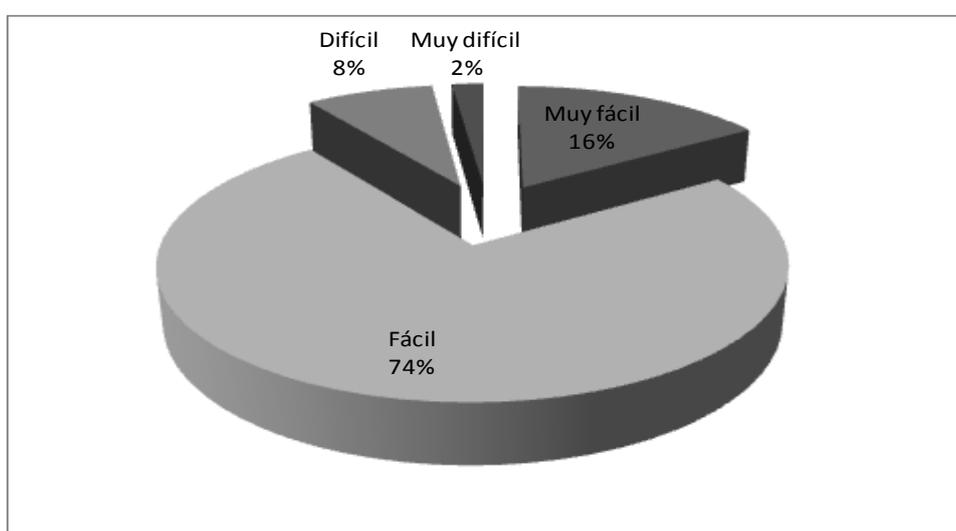


Fig.24. Dificultad de la búsqueda por temas

Les parece fácil siempre y cuando conozcan la interfaz de búsqueda y a partir de una palabra obtienen un conjunto de resultados. Sin embargo, señalan que la sencillez del proceso se contrapone al resultado, ya que los registros obtenidos no siempre les son útiles.

Al comparar la búsqueda por temas con la búsqueda por autor o título, la opinión de los usuarios es que prefieren las opciones últimas no porque el proceso sea más sencillo o difícil, sino porque estas opciones de búsqueda ofrecen mejores resultados que la búsqueda por materias. Se observa, nuevamente, que los usuarios prefieren el uso de otros campos.

Sobre el proceso de la búsqueda por temas, los usuarios destacan dos momentos difíciles:

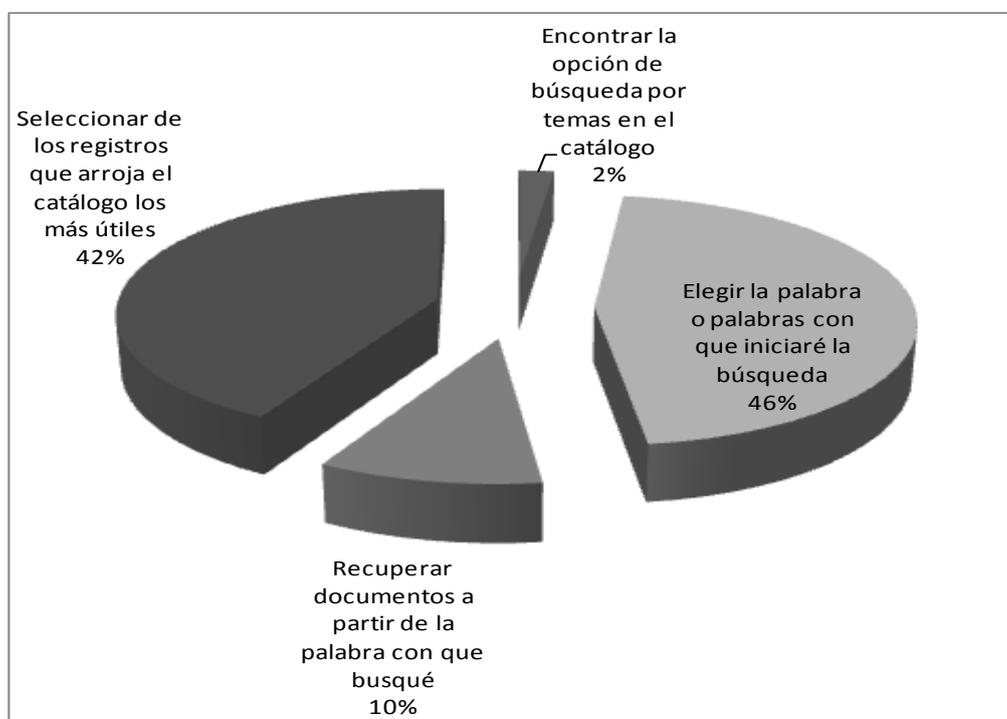


Fig. 25. Dificultad en el proceso de búsqueda por temas

Encontrar la opción de búsqueda por temas les resulta sencillo, incluso, comentan que la interfaz es amigable y fácil de usar porque se ha mantenido en las distintas versiones del catálogo. Elegir la palabra con la que iniciarán la búsqueda es lo más complicado, pues se preocupan por elegir la adecuada. Recuperar documentos a partir de la palabra con la que buscan no es complicado, pues casi siempre recuperan información. Mientras que seleccionar de los registros que arroja el catálogo los más útiles también es complicado ya que frecuentemente recuperan demasiados registros.

5.3. Formulación de la búsqueda

En las preguntas que indagan sobre la capacidad de los usuarios en la formulación de temas los resultados coinciden con los de las transacciones de temas. Se confirmó que los usuarios tienen graves problemas en la formulación de búsqueda. Son poco hábiles trasladando conceptos a símbolos comprensibles para el sistema de recuperación de información. Para la mayoría es sencillo dar una descripción del tema que necesitan, presento ejemplos: “Mi tema es sobre regionalismos, pero no relacionados con la política, sino con la literatura”, “Busco libros sobre el dialecto de lugares específicos”, “La perspectiva de género en el socialismo”, “Violencia sexual contra mujeres en marcos de conflicto armado”, “Atención de los servicios médicos en las víctimas que sufrieron violencia intrafamiliar”.

Su problema es la traducción del concepto en palabras de búsqueda. Al momento de formular la búsqueda emplean palabras demasiado generales. Así, para el primer tema el usuario buscaría por “Regionalismos”, “ Tradición” y “Literatura”; para el segundo tema el usuario emplearía las palabras “Socialismos”, “Mujer”, “Políticas públicas” y “Género”; para recuperar documentos sobre la violencia sexual contra mujeres en conflictos armados el usuario buscaría por “Justicia”, “Reparación”, “Conflicto armado” y “Violencia contra mujeres”; mientras que para el último tema sobre la atención de los servicios médicos a víctimas de violencia intrafamiliar el usuario optó por las palabras “Salud”, “Violencia” y “Médicos”. Usan palabras simples, muchas veces ambiguas. Al momento de transformar el concepto en palabras en lugar de ligar las palabras de manera que representen el tema lo que hacen es desfragmentarlo en término aislados que pierden el sentido del tema. Es por tal motivo que al introducir estas palabras en el catálogo recuperan demasiada información poco relacionada con el tema que tienen en mente o no recuperan documentos.

Esta idea sobre la dificultad que les implica la formulación de la búsqueda se comprueba, pues al preguntarles si esta etapa de la búsqueda temática es fácil o difícil se obtuvieron estas respuestas:

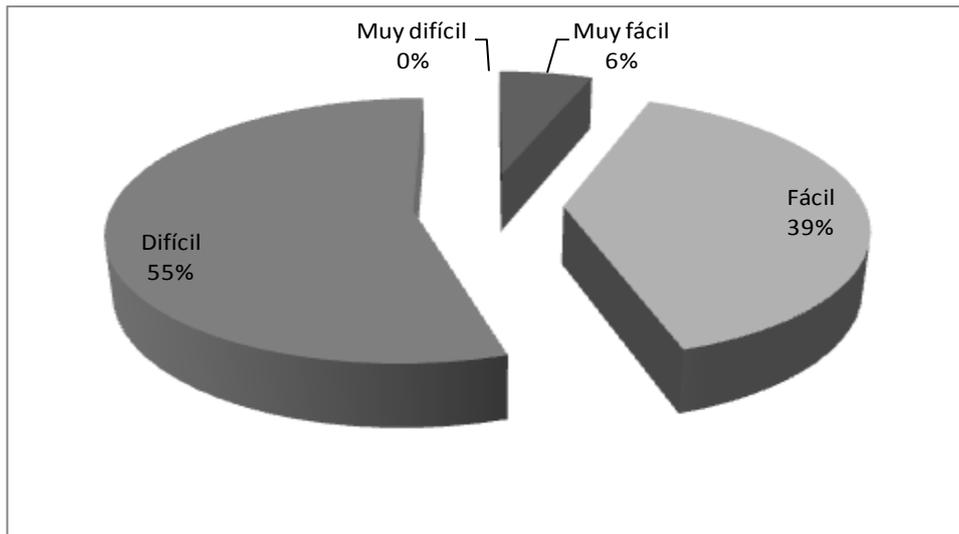


Fig. 26. Dificultad de la formulación de búsqueda

Para el 55% de los usuarios es difícil, mientras que para el 39% es fácil. Los usuarios para quienes es sencillo formular la búsqueda parecen conocer más la estructura de la búsqueda por temas, pues en sus formulaciones vinculan términos, por ejemplo, “Paul Valery- Crítica e interpretación”, “Poesía española-Siglo XX”, “Mujeres-Mortalidad”. La formulación de la búsqueda se da frente al catálogo, por lo que es un proceso de interacción con el sistema. Sobre el estado de su primera formulación al buscar un tema señalaron que:

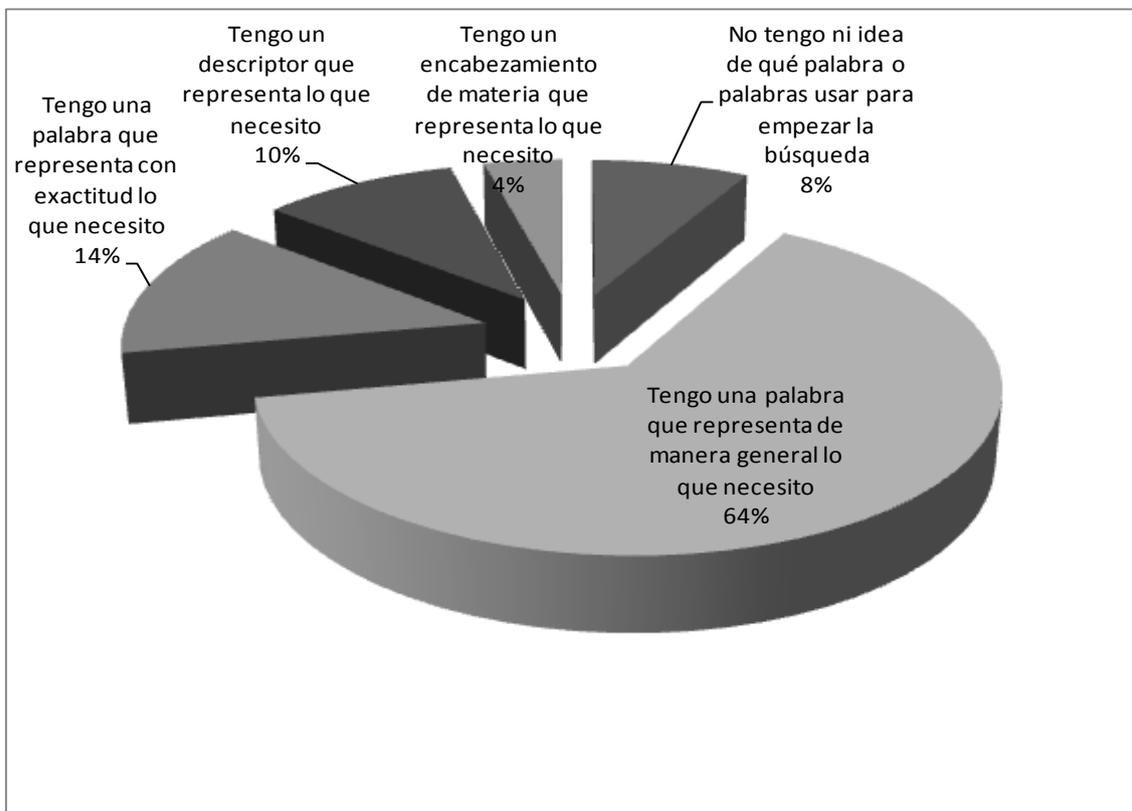


Fig. 27. Estado de la formulación de búsqueda frente al catálogo

El 64% antes de introducir las palabras de búsqueda tienen una palabra que representa de manera general lo que necesitan. Un 8% no tienen ni idea de qué palabra pueden usar para buscar. Una porción pequeña 14% saben tienen una palabra exacta que consideran representa lo que necesitan. Mientras que un 10% tiene un descriptor que saben se usa para describir lo que necesitan y sólo un 4% conoce algún encabezamiento de materia que saben es usado para representar lo que necesitan.

Al interrogarlos sobre la dificultad que implica elegir la palabra de búsqueda y la asertividad de ésta se encontró que:



Fig. 28. Dificultad para elegir la palabra de búsqueda

Para la mayoría, 64%, es sencillo elegir la palabra de búsqueda, pero no siempre es la adecuada desde el punto de vista de su efectividad. Sólo para el 14% es sencillo localizar una palabra adecuada que obtenga buenos resultados en la recuperación de documentos. Un 10% considera que es un proceso complicado pero encuentran la adecuada y un 12% considera que es complicado y, además, no siempre es la adecuada. Los datos muestran que los usuarios perciben cierta dificultad en el manejo de las palabras de búsqueda y el modo como éstas se enlazan para la recuperación de documentos. A la vez, son conscientes de que para buscar por temas es necesaria la selección de palabras adecuadas. Sin embargo, no se cuestionan el por qué de esta dificultad, es decir, no son conscientes de la presencia de lenguajes documentales en la

organización de la información temática, aún cuando algunos dicen que para buscar por temas deben pensar qué palabra usó la biblioteca para referirse al tema que ellos buscan.

Cuando en sus búsquedas por temas no obtienen resultados, los usuarios, como se observa en el siguiente gráfico, en su mayoría, 78%, piensan que la palabra de búsqueda no fue la adecuada, un 12% cree que hizo algo mal en el procedimiento de búsqueda y un 10% acepta la idea de que la biblioteca no tiene documentos sobre el tema:

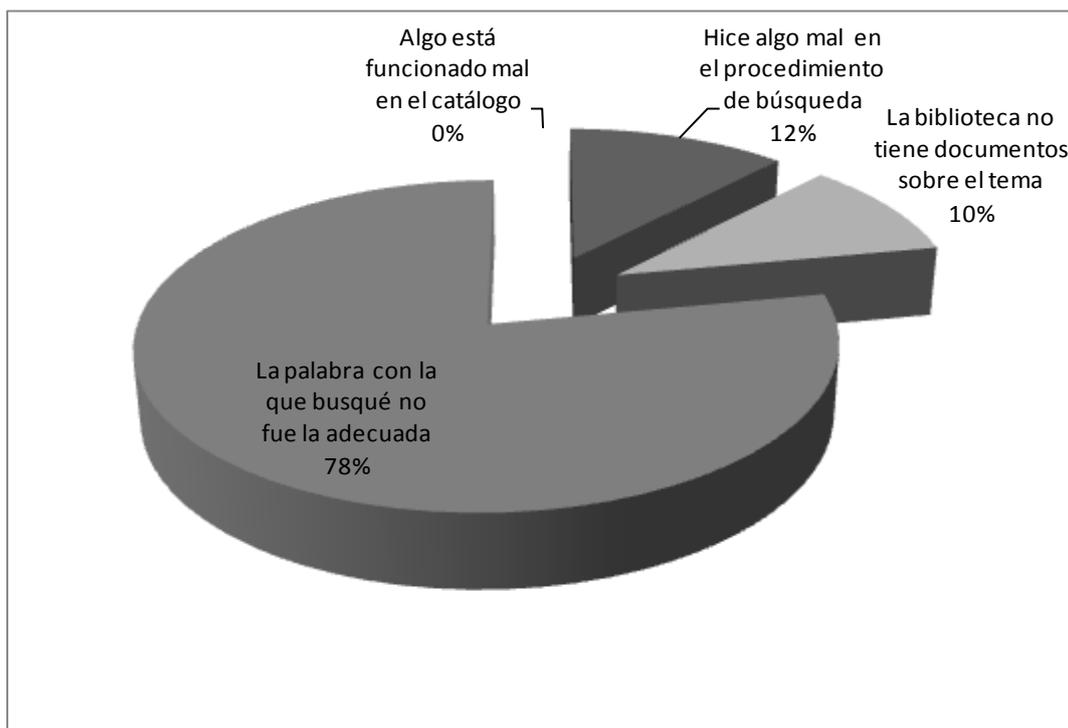


Fig. 29. Ideas del usuario cuando no recupera documentos

Una vez que los usuarios hacen su búsqueda y no obtienen resultados viene una segunda etapa donde se toma una nueva decisión. Al no obtener resultados, la opción más generalizada es buscar con otras palabras, sinónimos del tema, o buscar temas relacionados:

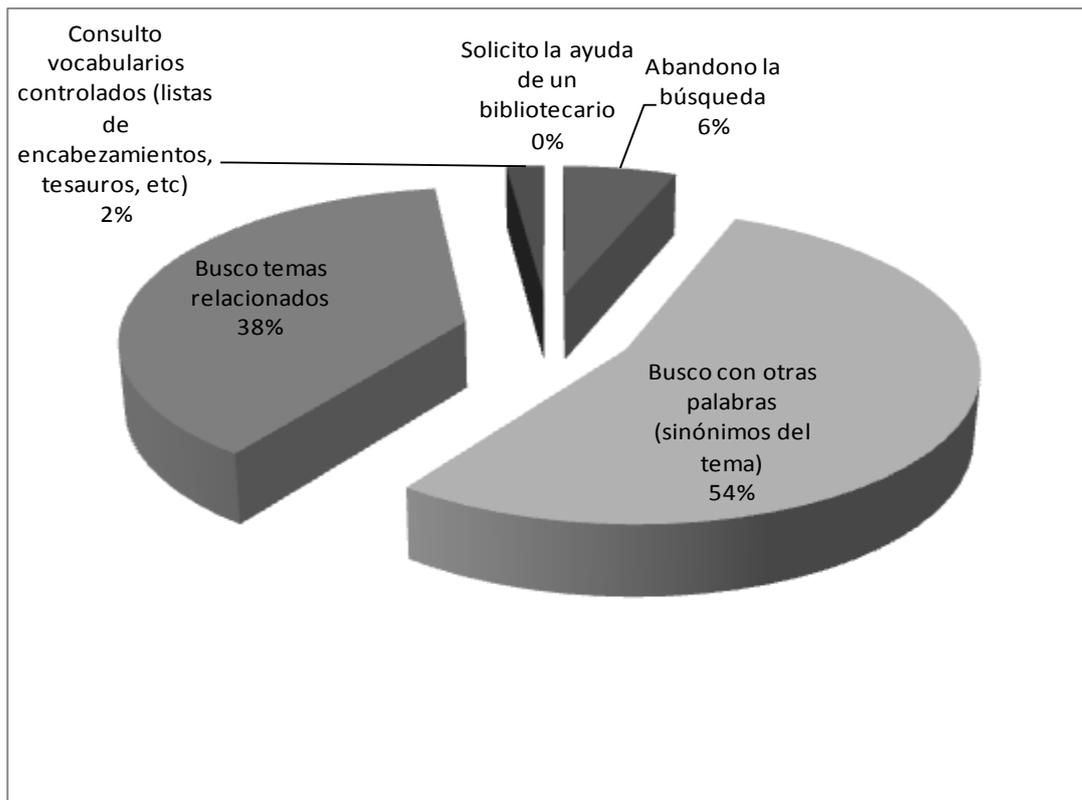


Fig. 30. Acciones del usuario cuando no recupera documentos

Los datos muestran que en la mayoría de los casos hay una reformulación de la búsqueda, ya sea a través de sinónimos del tema o temas relacionados. Sólo el 6% abandona la búsqueda. Y llama la atención que una cantidad muy pequeña, el 2%, consulta vocabularios controlados para dirigir con más precisión su búsqueda. Es sorprendente que ninguno de los encuestados considera que para dirigir mejor sus búsquedas por temas puede ser necesaria la ayuda de un bibliotecario.

5.4. Petición de búsqueda

Los bibliotecólogos pocas veces nos detenemos a pensar cómo busca el usuario. Este hecho resultó de interés para la presente tesis, es por ello que tanto en las transacciones de búsqueda por temas como en el cuestionario se estableció la pregunta. Los resultados obtenidos en el cuestionario sobre qué tipo de palabras usan en su búsqueda fueron los siguientes:

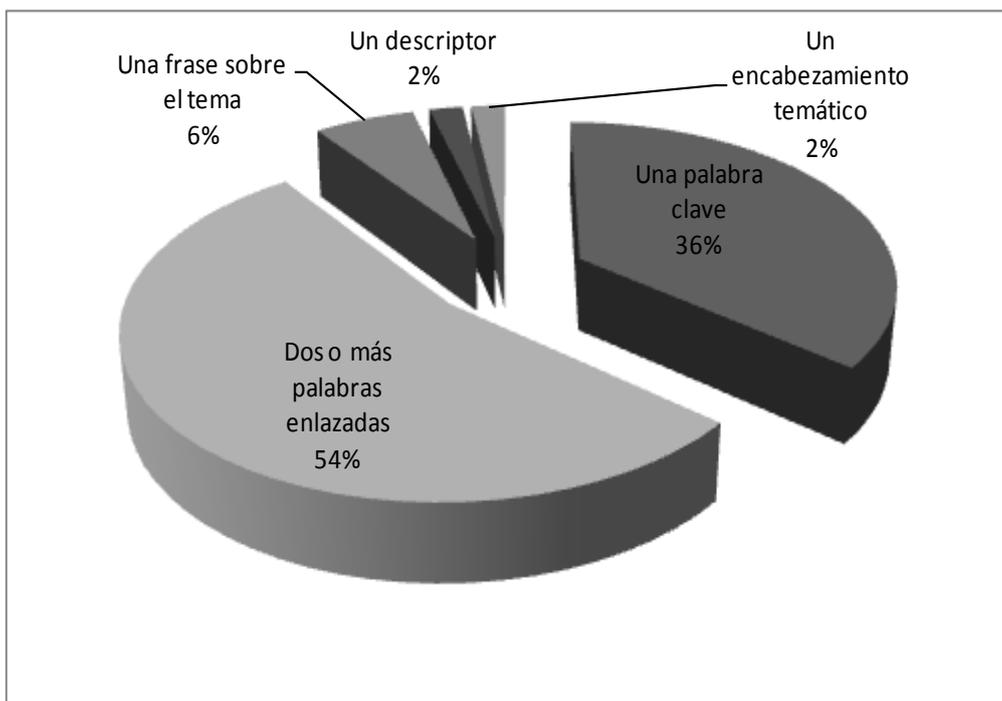


Fig. 31. Petición de búsqueda

Según estos datos, la estructura de más uso estaría formada por dos o más palabras enlazadas; sin embargo, no es así, pues el análisis de transacciones de temas, muestra que su búsqueda es más bien mediante palabras simples. Los datos de las transacciones temáticas son más confiables en tanto que se tiene la petición de búsqueda tal y como el usuario la escribió en el catálogo. Aunado a esto, al solicitar a los usuarios que transformaran su necesidad de información en términos de búsqueda, casi todos usaron palabras simples, aun cuando había dicho que en la búsqueda siempre optaban por usar dos o más palabras enlazadas.

5.5. Conocimientos del usuario sobre el lenguaje de indización

Los usuarios de la BDCV en su mayoría desconocen tanto el uso del lenguaje controlado en la indización como la estructura de estos. Al preguntarle a los usuarios si identifican la parte del registro donde parecen los temas que la biblioteca asigna a los documentos. Mientras que en la biblioteca la catalogación temática se lleva a cabo con meticulosidad, la actividad no se vincula con el área de servicios al público ni con la formación de usuarios. La mayoría de los usuarios desconocen la existencia de temas asignados a los materiales de la biblioteca, el cuestionario mostró que:

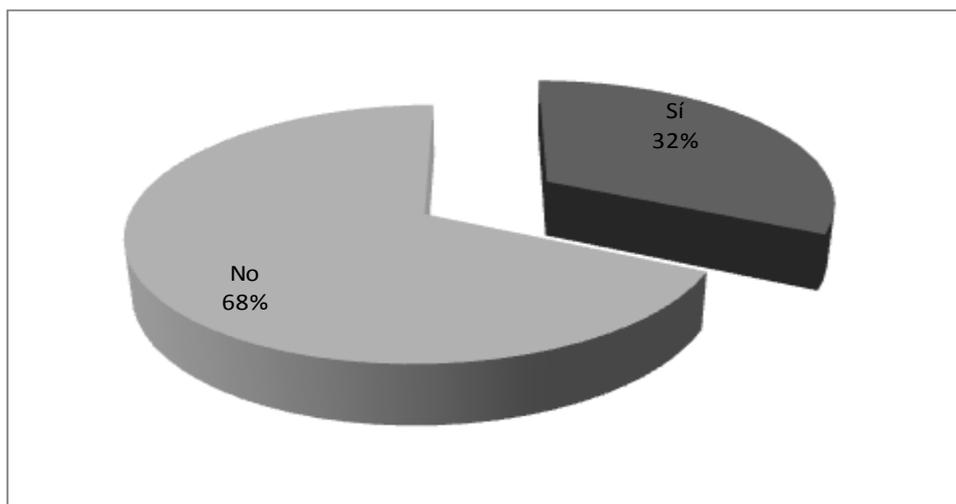


Fig. 32. Conocimiento de la indización

Sólo el 32% de los usuarios sabe en qué parte del registro se pueden ver los temas que la biblioteca asigna a los documentos. El resto, 68% no sabe. Muchos de ellos, incluso, se mostraron sorprendidos, pues no se habían dado cuenta de que en la parte baja de los registros, bajo la etiqueta temas aparecen varios encabezamientos que pueden redireccionarse como búsquedas temáticas.

Al 32% de los usuarios que sabían de la presencia de los encabezamientos temáticos en el registro, se les preguntó si los revisan en sus búsquedas por temas, se encontró que:

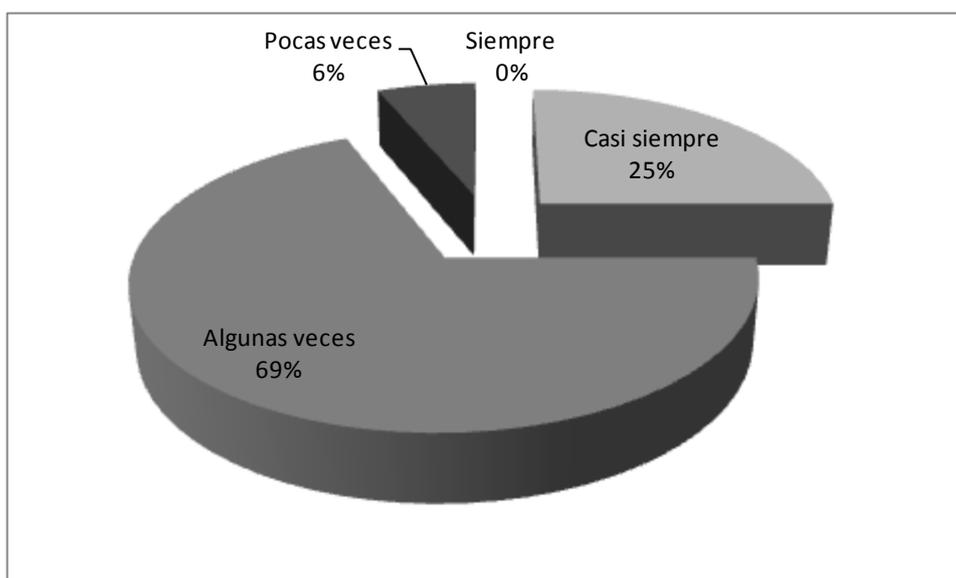


Fig. 33. Uso de los encabezamientos temáticos asignados en la indización

El 6% los revisa casi siempre para re direccionar sus búsquedas por temas. El 69% los usa algunas veces, cuando considera que alguno le resulta útil y un 2% los

revisa pocas veces, pues no le parecen útiles para recuperar más documentos y opta por introducir otras palabras de búsqueda diferentes a las opciones de los encabezamientos.

Los usuarios de la biblioteca no se percatan del enlace o falta de éste entre los términos con los que buscan y los temas asignados. Al interrogarlos sobre si las palabras que introducen coinciden con los temas que asigna la biblioteca sus respuestas fueron:

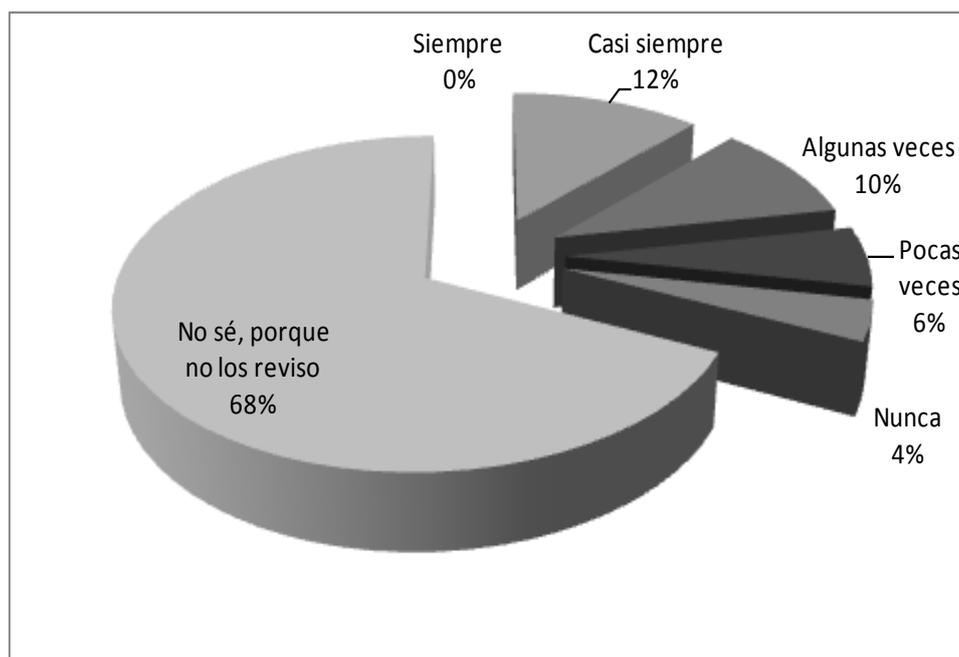


Fig.34. Coincidencia entre la petición e indización

Un 12% dice que sus peticiones se enlazan a la indización casi siempre. El 10% mencionó que algunas veces. Pocas veces obtuvo el 6% y un 4% dijo que sus peticiones nunca coinciden con los temas de la biblioteca. El resto, 68%, no sabe si su petición se enlaza o no con la indización dado que no revisa los temas asignados por la biblioteca.

5.6. Recuperación en la búsqueda

La recuperación de documentos en la biblioteca es una cifra simbólica, en tanto que un usuario puede obtener gran cantidad de registros, pero éstos no siempre son útiles a su necesidad de información. La impresión de quienes usan el catálogo es importante en este aspecto pues en la recuperación ellos se quedan con una impresión de la eficiencia del catálogo y la capacidad de éste para vincular con las colecciones de la biblioteca. En cuanto a la precisión de la recuperación los resultados fueron los siguientes:

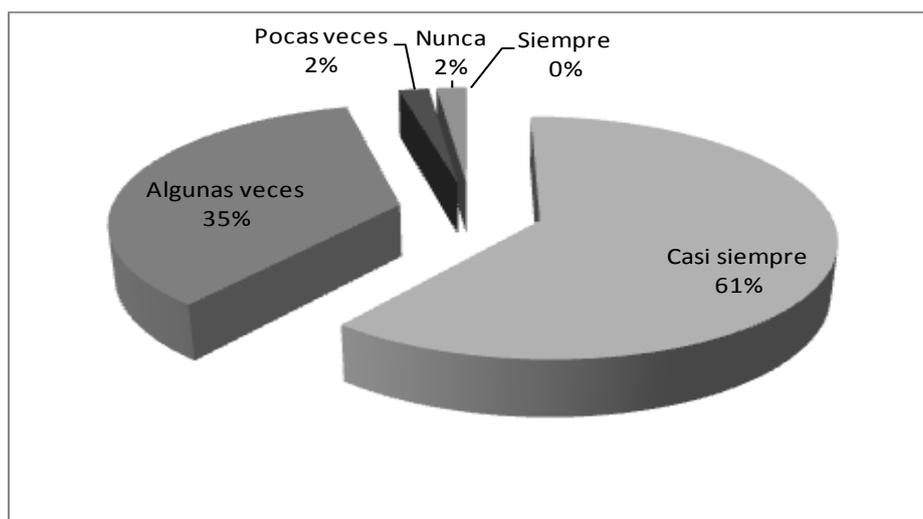


Fig. 35. Recuperación en la búsqueda

El 61% considera que recupera documentos útiles casi siempre, mientras que el 35% sólo algunas veces, el 2% pocas veces y un 2% nunca recupera materiales que satisfacen su necesidad de información. Estos datos coinciden con las cifras de recuperación que se obtuvieron en las transacciones de temas donde el 66% recuperaba documentos y el 33% no recuperaba. Lo que bien nos puede llevar a inferir que en el porcentaje que recupera documentos uno o varios de éstos son útiles al usuario.

Con la pregunta sobre si los documentos recuperados tratan el tema que el usuario buscaba se comprueba lo enunciado anteriormente:

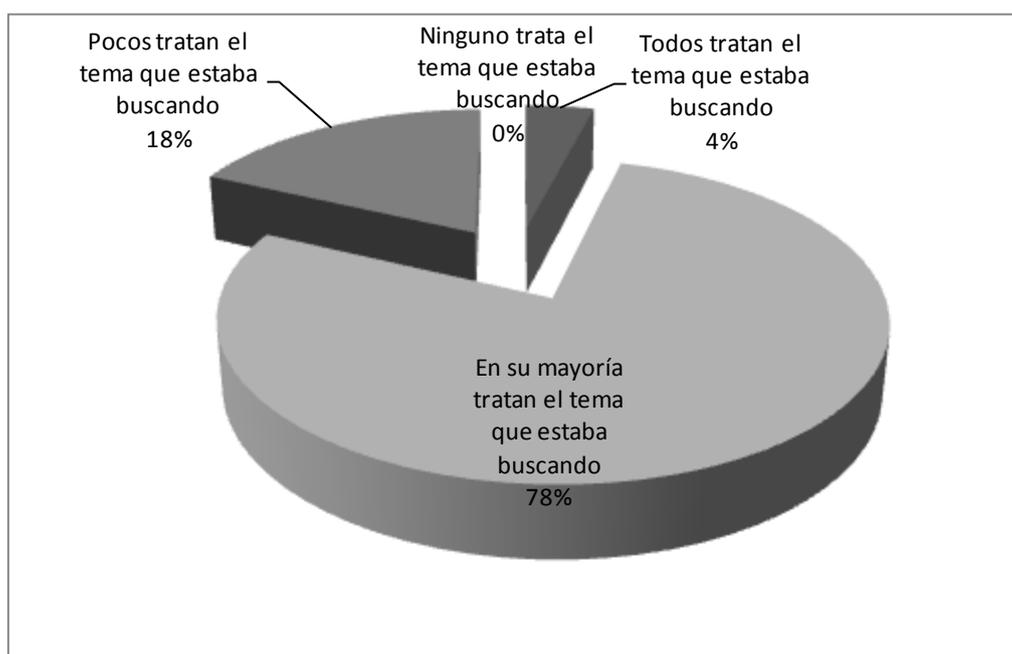


Fig.36. Utilidad de los documentos recuperados

La impresión más generalizada es la mayoría de los documentos tratan el tema que estaban buscando. El 18% dice que los materiales no trataban el tema en cuestión. Esta cifra tal vez se deba a peticiones en la que hay recuperación pero esta es desfragmentada, es decir, esas peticiones de las que se habló en el apartado de transacciones temáticas en las que el catálogo no puede enlazar la petición completa y la divide, buscando por palabras clave en los encabezamientos, lo que crea una descontextualización de la solicitud.

Los usuarios consideran que en la búsqueda por temas tienen problemas para recuperar documentos. En algunas ocasiones estos problemas se asocian con características del tema buscado:

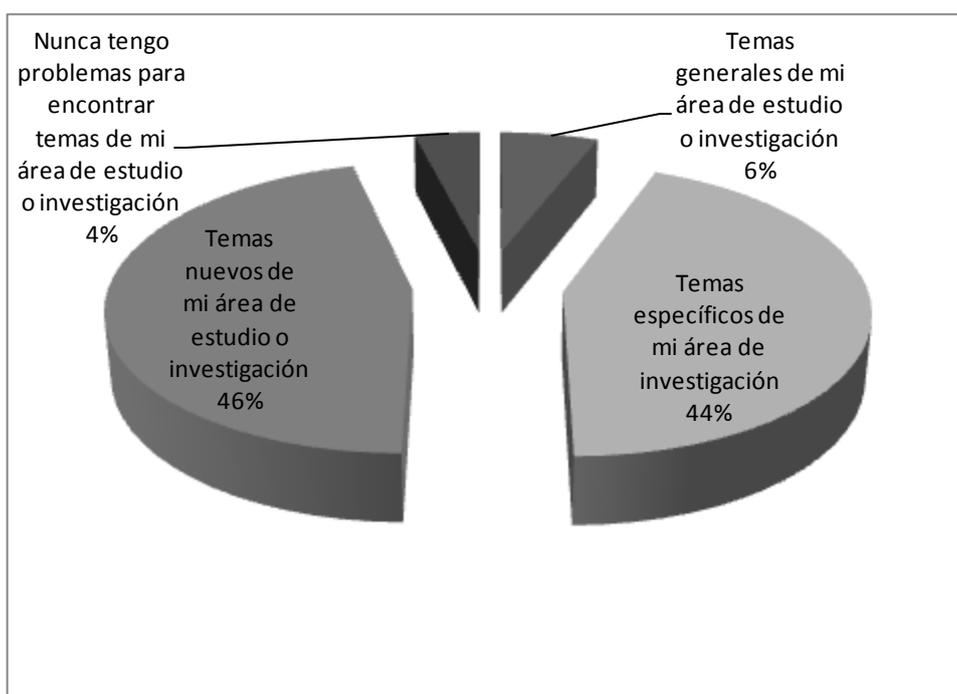


Fig. 37. Problemas en la recuperación relacionados con características del tema

Destacan que cuando buscan temas generales casi no tienen problemas. Éstos se presentan cuando buscan temas específicos o temas nuevos de su área de investigación. Entre los comentarios que agregan es que consideran que muchos temas nuevos o que son muy específicos de un área la biblioteca no los incluye. Esto se da porque las herramientas de indización tardan en actualizarse en relación con la terminología de las disciplinas que están en continuo movimiento, mientras que los temas específicos, debido a la limitación jerárquica que imponen los sistemas de organización del conocimiento a veces suelen estar incorporados dentro de categorías superiores.

4.7. Sugerencias de los usuarios para mejorar la búsqueda

Las sugerencias de los usuarios para mejorar la búsqueda por tema se agrupan en cuatro categorías:

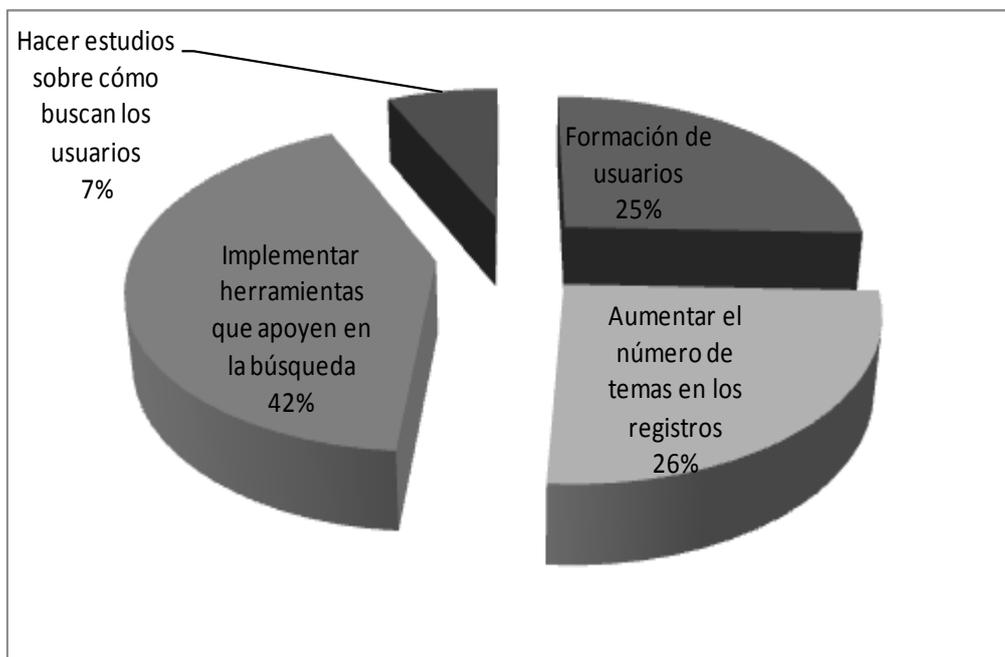


Fig.38. Sugerencias de los usuarios para mejorar la búsqueda por temas

La petición más frecuente es que la biblioteca les brinde herramientas que los auxilien en sus búsquedas por temas. Sugieren la construcción de una herramienta que los ayude para expresar mejor el tema, no saben que la biblioteca cuenta con índices de temas y que se puede acceder a ellos mediante esa búsqueda particular. Poner en un espacio más visible esta herramienta puede ayudarles para mejorar sus búsquedas. Otra acción para ayudarlos es formarlos para que sean capaces de formular búsquedas complejas en las que recuperen información precisa. La tercera sugerencia que dan: hacer estudios sobre cómo buscan los usuarios también resulta necesario, pues en la biblioteca no hay estudios al respecto.

4.8. Conclusión

Mediante los resultados obtenidos del cuestionario se determinó que la búsqueda por temas es poco usada en la BDCV, ya que sólo es empleada cuando el usuario no tiene el dato preciso a buscar y salvo en un primer momento de la investigación relacionado con la búsqueda de bibliografía. Quienes buscan por temas coinciden al señalar que el punto crítico de ésta es la formulación, pero este hecho no la hace más difícil que otras

opciones para recuperar información. La comunidad de El Colegio de México parece estar satisfecha con los resultados que obtiene en su búsqueda por temas, pues consideran que los documentos recuperados le son útiles y tratan el tema buscado. Su opinión es que sólo se enfrentan a problemas cuando buscan temas específicos o nuevos de su área de investigación.

En lo que respecta al encuentro entre sus peticiones y la indización, los usuarios no son muy conscientes de que en esta última se emplean lenguajes controlados. Un 68% no sabe que existen, ni sabe que los registros bibliográficos contienen temas asignados a los documentos. Mientras que del 32% que sabe de su presencia el 28% ha revisado en alguna ocasión los temas y ha visto si éstos tienen relación con las palabras que empleó en la búsqueda. El trabajo de indización es casi desconocido para los usuarios y, si bien cierto que no están obligados a saberlo, si es necesario que lo conozcan pues este conocimiento es un factor importante en la búsquedas por temas.

Finalmente, este trabajo detectó que la búsqueda por temas en la BDCV es poco usada. El hecho podría verse como alarmante y llevaría a pensar que todo el trabajo de indización no está siendo útil. Sin embargo no es así, los usuario de la biblioteca no la emplean mucho debido a que prefieren buscar a partir de datos precisos. Para ellos la búsqueda por temas no implica dificultad y los resultados obtenidos en ésta les resultan aceptables en función de su utilidad. En tal sentido, seguir con la actividad y más aún reforzarla significa seguir brindando a los usuarios un campo de búsqueda que les será útil cuando necesiten buscar por temas.

Conclusiones generales

El encuentro del lenguaje de indización con el lenguaje de búsqueda en la recuperación por temas presenta dos problemas vinculados al uso del lenguaje controlado en la indización y el uso del lenguaje natural en la búsqueda:

- a) Formulación de búsqueda: se observa que el usuario desconoce, en general, los principios del lenguaje controlado y en el catálogo pretende buscar como en un buscador de Internet donde mediante el uso de palabras del lenguaje natural es posible recuperar información. En las formulaciones de búsqueda pocos usuarios usan estructuras compuestas, operadores booleanos y menos aún facetas para delimitar la recuperación. Suelen emplear palabras simples y generales bajo las cuales obtienen muchos documentos y en los que tienen que invertir gran cantidad de tiempo discriminando su utilidad. Otras ocasiones la formulación no se concreta en símbolos comprensibles para el sistema, es el caso de frases donde no hay recuperación o peticiones en extremo ambiguas que en la recuperación obtienen resultados dispersos y diferentes conceptualmente a la solicitud planteada.
- b) Enlace de la petición con la indización : la estructura controlada usada en la organización de la información resulta poco flexible debido a que no hay herramientas que auxilien en el proceso (índices, autoridades visibles, etc.) y ayuden para que el usuario enlace sus peticiones a los términos de indización.

Los problemas de encuentro entre la petición se deben en gran medida al alejamiento entre ambos lenguajes y los principios que los rigen en tres niveles: semántico, sintáctico y léxico.

Bajo el principio semántico el lenguaje controlado busca eliminar en lo posible la sinonimia, polisemia y homonimia. Su principio es controlar los fenómenos a través del uso de referencias cruzadas. Esto es útil, pero surge una pregunta ¿es posible controlar estos fenómenos? Es muy probable que no, porque hasta ahora la disciplina que más ha trabajado al respecto: la terminología, sigue lidiando con ellos. Por otra parte, en la búsqueda por temas se observa que las solicitudes están cargadas de polisemia y sinonimia.

Bajo el principio de sintaxis el lenguaje controlado emplea una sintaxis regida por la facetación. Es así que un documento se indizará bajo un tema principal y para delimitarse se le agregarán facetas según una sintaxis establecida. Quienes somos bibliotecarios conocemos este orden, pero el usuario no. Él no es catalogador, no conoce sobre el tema y busca con la sintaxis de su lengua. Lo que le implica graves problemas cuando intenta recuperar información por temas.

Bajo el principio léxico, los vocabularios controlados tienen un conjunto de términos aceptados que excluyen muchas manifestaciones léxicas empleadas por los usuarios, es el caso de regionalismos, jerga e incluso tecnicismos especializados que los vocabularios controlados no contemplan como términos aceptados y que los registros de autoridades no integran.

Se concluye que el lenguaje de búsqueda es muy distinto al lenguaje de indización. Pero, lo importante es qué pueden hacer al respecto. La primera sugerencia es aplicar más el principio de uso de los vocabularios controlados, interesémonos por saber cómo buscan nuestros usuarios. Segundo, apostemos por su formación, la investigación detectó que en la BDCV hay usuarios muy hábiles en la formulación de búsquedas que obtienen resultados muy satisfactorios. Tercero, retroalimentemos nuestras actividades con la investigación y la evaluación.

Gran parte de la literatura sobre recuperación de información señala que los usuarios en un sistema de recuperación deben plantear su necesidad en términos comprensibles para el sistema. Pero, el paradigma centrado en el usuario se opone a esta afirmación: ¿es el usuario quien debe adecuarse a la estructura del sistema o viceversa? Si se quiere continuar haciendo el trabajo como hasta ahora, repitamos lo mismo: que el usuario se adapte a sistema. Sin embargo, en un contexto donde los buscadores de Internet y las bases bibliográficas brindan mayor flexibilidad en la búsqueda, la indización debe centrarse más en el usuario. Es por eso que la biblioteca además de enfocarse en la organización de los materiales debe vincular esta actividad con el diseño de herramientas más acordes con las necesidades de los usuarios y debe instruirlos para que obtengan mejores resultados en sus experiencias de búsqueda.

Otro aspecto fundamental es el desarrollo de la investigación en bibliotecología. La tarea de los profesionales en bibliotecología y de los egresados de la Maestría en bibliotecología de El Colegio de México es llevar a cabo investigaciones de calidad que contribuyan al enriquecimiento de la disciplina y el mejor funcionamiento de las bibliotecas del país.

Esta tesis es un primer estudio sobre la recuperación temática en catálogos de bibliotecas, específicamente en la Biblioteca Daniel Cosío Villegas. De modo preciso, ha cumplido los objetivos que se planteo; no obstante, es un trabajo que abre un gran abanico de futuras investigaciones que los bibliotecarios egresados de El Colegio de México y las distintas escuelas de bibliotecología del país debemos llevar a cabo.

Bibliografía

- Adams, J.A. "The computer catalog: a democratic or authoritarian technology". En *Library journal*. Nueva York: R. R. Bowker, vol. 113, núm. 2, 1988. Pp. 31-36.
- Aluri, Rao, Alasdair D Kemp y John J. Boll. *Subject analysis in online catalogs*. Englewood, Colorado: Libraries Unlimited, 1991.
- Antell, Karen y Jie Huang. "Subject searching success: transaction logs; patrons perceptions and implications for library instruction". En *Reference and user services quarterly*. Chicago: American Library Association, vol. 48, núm. 1, 2009. Pp. 68-76.
- Arriola Navarrete, Oscar. *Documentos básicos para la catalogación de materiales en la Biblioteca Daniel Cosío Villegas*. México: El Colegio de México, Biblioteca Daniel Cosío Villegas, 1997.
- Barite Roqueta, Mario Guido. "Los conceptos y su representación: una perspectiva terminológica para el tratamiento temático de la información". En *Scire: representación y organización del conocimiento*. Zaragoza, España: Universidad de Zaragoza, vol. 2, núm.6, 2000. Pp. 31-53.
- Bates, Marcia. "Subject access in online catalogs: a design model". En *Journal of the american society for information science*. Nueva York: John Wiley & Sons, vol. 37, núm. 6, Noviembre, 1986. Pp. 357-376.
- _____. "System meets user: problems in matching subject search terms". En *Information processing and management*. Oxford: Pergamon, vol. 13, núm. 6, 1977. Pp. 367-375.
- Beall, Jeffrey. "Search Fatigue". En *American libraries*. Chicago, Ill.: American Library Association, vol. 38, núm. 3, 2007. Pp. 46-50.
- Belkin, N. J. "Anomalous state of Knowledge as a basis for information retrieval". En *Canadian Journal of Information Science*. Toronto: Canadian Association for Information Science, vol. 5, núm. 3, 1980. Pp. 133-143.
- Belkin, N. J., R. N. Oddy y H. M Brooks. "ASK for information retrieval: part 1. background and theory". En *Journal of documentation*. Londres: Association of Special Libraries and Information Bureaux, vol. 38, núm. 2, 1982. Pp. 61-71.
- Biblioteca Daniel Cosío Villegas, Coordinación de acceso a bases de datos. *Políticas de catalogación temática*. México: Biblioteca Daniel Cosío Villegas, Coordinación de acceso a bases de datos, c2010. (DE, 5 de noviembre, 2010: <http://biblio.colmex.mx/Intranet/corabd.politicascatalog.html#autoridades>).
- Biblioteca Luis Angel Arango. *Lista de encabezamientos de materia para bibliotecas*. Santafé de Bogotá: R. Eberhard, 1998.
- Bonura Larry. *The art of indexing*. Nueva York: J. Wiley & Sons, 1994.

- Borgman, Christine L., Dineh Monghdam y Patti K. Corbett. *Effective online searching*. Nueva York: Marcel Dekker: New York and Basel, 1984.
- Borgman, Christine. "Why are online catalogs still hard to use". En *Journal of the american society for information science*. Nueva York: John Wiley & Sons, vol. 47, núm. 7, 1996. Pp. 493-503.
- Brice, Allen. "Topic knowledge and online catalogue search formulation". En *Library quarterly*. Chicago: Universidad de Chicago, vol. 61, núm. 2. Pp. 188-213.
- Carlyle, Allyson. "Matching LCSH and user vocabulary in the library catalog". En *Cataloging and classification quarterly*. Nueva York: Haworth, vol. 10, núm.1, 1989. Pp. 37-63.
- Chowdhury, G.G. *Introduction to modern information retrieval*. Londres: Facet Publishing, 2004.
- Chu, Heting. *Information, representation and retrieval in the digital age*. Meford, NJ.: Information today, 2003.
- Cousins, Shirley Anne. "Enhancing subject access to OPACS: controlled vocabulary vs natural language". En *Journal of documentation*. Londres: Association of Special Libraries and Information Bureaux, vol. 48, no.3, 1992. Pp. 291-309.
- Dervin, B. "Information as user construct the relevance of perceived information needs to synthesis and interpretation". En Ward, S.A y Reed J. L., eds. *Knowledge structures and use. Implications for synthesis and interpretation*. Filadelfia: Temple University Press, 1983. Pp. 153-184.
- Escamilla González, Gloria. *Lista de encabezamientos de materia*. México: UNAM, Instituto de Investigaciones Bibliográficas, 1978.
- España. Consejo Superior de Investigaciones Científicas. Unidad de Coordinación de Bibliotecas. *Lista de encabezamientos de materia de la Red de Bibliotecas del CSIC*. Madrid: Consejo Superior de Investigaciones Científicas, 2000.
- El Colegio de México. *Portal de El Colegio de México*. México: El Colegio de México, 2010. (DE, 31 de Agosto, 2010: <http://www.colmex.mx>).
- Feather, John y Paul Sturges. *International encyclopedia of information and library science*. Nueva York: Routledge, 1997.
- Fernandez Molina, Juan Carlos y Felix de Moya Anegón. *Los catálogos de acceso público en línea: el futuro de la recuperación de información bibliográfica*. Granada: Asociación andaluza de bibliotecarios, 1998.
- Figueroa-Servin, Reynaldo D. y Berta Enciso. "Subject authority control at El Colegio de Mexico's library: the whats and hows of a project." En *Cataloging & classification quarterly*. Nueva York: Haworth, vol. 32, núm. 1, 2001. Pp. 65-80.

- Foskett, A. C. *The subject approach to information*. Hamden, CT.: Linnet Books: Clive Bingley, 1972.
- Graham, Rumi Y. "Subject no-hits searches in an academic library online catalog: an exploration of two potential ameliorations". En *College and research libraries*. Chicago: American Library Association, vol. 65, num.1, 2004. Pp. 36-54.
- Gross, Tina y Arlene G.Taylor. "What have we got to lose? The effect of controled vocabulary on keyword searching results". En *College and research libraries*. Chicago: American Library Association, vol.66, núm.3, 2005. Pp. 212-230.
- Hancock-Beaulieu, Micheline. "Evaluating the impact of an online library catalogue and subject searching behaviour at the catalogue and at the shelves". En *Journal of documentation*. Londres: Association of Special Libraries and Information Bureaux, vol. 46, núm.4, 1990. Pp. 318-338.
- Hert, Carol A. "User goals on an online public access catalog". En *Journal of the american society for information science*. Nueva York: John Wiley & Sons, vol. 47, núm. 7, 1996. Pp. 504-518.
- Hildreth, Charles R. *The online catalogue: developments and directions*. London: The Library Association, 1989.
- Hjørland, Birger. *Information seeking and subject representation*. Westport, CT.: Greenwood Press, 1997.
- Hunter, R. N. "Succeses and failures of patrons searching the online catalog a large academic library: a transaction log analysis". En *Research Quarterly*. Alexandría, VA.: SHRM Research, vol. 30, núm. 4, 1991. Pp. 395-402.
- Hutchins, J.W. "The concept of 'aboutness' in subject indexing". En *Aslib Proceedings*. Londres: ASLIB, vol. 30, núm. 5, 1978. Pp.172-181.
- _____. "On the problem of 'aboutness' in document analysis". En *Journal of Informatics*. Statesboro, GA. : International Academy for Information Management, vol.1, núm.1, 1977. Pp. 17-35.
- Ingwersen, Peter. *Information retrieval interaction*. Los Ángeles, CA.: Taylor Graham, 1992.
- _____. "Cognitive Perspectives of Document Representation". En *CoLIS 4 : fourth international conference on conceptions of library and information science : emerging frameworks and methods : (Seattle WA, 21-25 July 2002)*. Seattle, WA.: Universidad de Washintong, 2002. Pp. 285- 300.
- Izquierdo Arroyo, José María y Luis Miguel Moreno Fernández. "Listas de encabezamientos de materia y thesauri en perspectiva comparada". En *Documentación en ciencias de la información*. Madrid: Universidad Complutense, núm. 17, 1994. Pp.287-310.

- Kaske, N. K. "A comparative study of subject searching in an OPAC among branch libraries of a university library system". En *Information technology and libraries*. Chicago, Ill.: Library and Information Technology Association, vol. 7, núm. 4, 1988. Pp. 359-372.
- Kent, Allen, Harold Lancour y Jay E. Daily, eds. *Encyclopedia of library and information science*. Nueva York: Marcel Dekker, 1974.
- Knutson, Gunnar S. "Subject enhancement: report on an experiment". En *College and research libraries*. Chicago: American Library Association, vol. 52, 1991. Pp. 65-79.
- Tillie Krieger. *Instructor influences versus text influences in the selection of subject descriptors by undergraduate students*. Tesis doctoral: Universidad de Illinois, Urbana-Champaign, 1981.
- Lancaster, F. W. *Vocabulary control for information retrieval*. Arlington, VA.: Information Resources Press, 1986.
- Lancaster, F. W, et. al. "Identifying barriers to effective subject access in library catalogs". En *Library resources and technical services*. Chicago: American Library Association, vol 35, núm. 4, 1991. Pp. 377-391.
- Larson, Ray. "The decline of subject searching: long-term trends and patterns of index use in an online catalog". En *Journal of the american society for information science*. Nueva York: John Wiley & Sons, vol. 42, núm. 3, 1991. Pp. 197-215.
- Lerat, Pierre. *Las lenguas especializadas*. Barcelona: Ariel, 1997.
- Lester, Marilyn Ann. *Coincidence of user vocabulary and library of congress subject headings: experiments to improve subject access in academic library online catalogs*. Tesis doctoral. Universidad de Illinois at Urbana-Champaign, 1989.
- Lopez Yepes, José, ed. *Diccionario enciclopédico de ciencias de la documentación*. Madrid: Síntesis, 2004.
- Marcos, Mari Carmen. *Mejoras en la consulta y presentación de los resultados en catálogos de bibliotecas*. Barcelona: Sección Científica de Ciencias de la Documentación., Departamento de Ciencias Políticas y Sociales, 2003. (DE, 13 de septiembre, 2009: www.mcmarcos.com/pdf/2003_mejoras-vigo.pdf)
- Markey, Karen. "Users and the online catalog subject access problems". En Matthews, Joseph R., ed. *The impact of online catalogs*. Nueva York: Neal Schuman, 1986.
- _____. *Online catalog use: results of surveys and focus group interviews in several libraries*. Washintong: OCLC, 2000. (DE, 13 de Agosto, 2010: <http://library.oclc.org>).
- Moreno Jiménez, Pilar. *Lenguajes de indización libres o controlados: la experiencia de la Biblioteca Daniel Cosío Villegas de El Colegio de México*. XII Coloquio sobre

Investigaciones Bibliotecológicas: UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 7-8 de Julio de 1994. Mesa redonda Lenguajes de Indización Libres o Controlados.

O' Brien, A. "Online catalogs: enhancements and developments". En Willams , M. E., ed. *Annual review of information science and technology*. Medford, NJ. : Learned Information, vol.29, 1994. Pp. 219-242.

Ogden, Charles Kay y I.A. Richards. *El significado del significado: una investigación acerca de la influencia del lenguaje sobre el pensamiento y de la ciencia simbólica*. Buenos Aires: Paidós, 1954.

Peters, Thomas A. "The history and development of transaction log analysis". En *Library hi tech*. Ann Arbor, Mich.: Pierian, vol. 11, núm. 2, 1993. Pp. 41-66.

Programa Cooperativo de Control de Autoridades (PCCA), Comité Técnico. *Políticas para la elaboración de autoridades de materia*. México: El Comité, 2009.

Reitz, Joan M. *Dictionary for library and information science*. Westport, CT.: Libraries Unlimited, 2004.

_____. *ODLIS: Online dictionary for library and information science*. Westport, CT: Libraries Unlimited, 2004-2010. (DE, 20 de Agosto, 2010: <http://lu.com/odlis/>).

Roe, Sandy. "Online subject access". En *Journal of internet cataloging*. Binghamton, N.Y.: Haworth, vol. 2, núm.1, 1999. Pp. 69-78.

Rowley, Jennifer. "The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research". En *Journal of information science*. Los Ángeles: SAGE, vol. 20, núm. 2, 1994. Pp. 108-119.

Sánchez Zurita, Juan Manuel. *Los usuarios Google frente al catálogo bibliotecario de la UNAM*. México: UNAM, Dirección General de Bibliotecas, 2009. (DE, 30 de Septiembre, 2009: <http://juanzurita.wordpress.com>).

Sheble, Lallura y Barbara M. Wildemuth. "Transactions logs". En Bárbara M. Wildemuth, ed. *Applications of social research methods to questions in information and library science*. Westport, Conn. : Libraries Unlimited, 2009

Sauperl, Alenka. *Subject determination during the cataloging process*. Lanham, Md.: London: Scarecrow, 2002.

Spink, A y T. Saracevic."Interaction in information retrieval: selection and effectiveness of search terms". En *Journal of the american society for information science*. Nueva York: John Wiley & Sons, vol. 48, núm. 8, 1998. Pp. 741-761.

Spink, Amanda y Charles Cole, eds. *New directions in cognitive information retrieval*. Dordrecht, Netherlands: Springer, 2005.

- Sridhar, M. S. "Subject searching in the OPAC of a special library: problems and issues". En *OCLC systems and services: international digital library perspectives*. Westport, CT. : Meckler, vol. 20, núm. 4, 2004. Pp. 183-191.
- Svenonius, Elaine. "The epistemological foundations of knowledge representations". En *Library trends*. Urbana, Ill. : University of Illinois Library School, vol. 52, núm. 3, 2004. Pp. 571-587.
- Taylor, Robert. "Question negotiation and information seeking in libraries". En *College and research libraries*. Chicago: American Library Association, vol. 29, núm.3, 1968. Pp. 178-194.
- _____. "The process of asking questions". En *American Documentation*. Washington, D.C. : American Documentation Institute, vol. 36, núm. 3, 1962. pp. 391-397.
- Tolle, John E. *Current utilization of online catalogs: transaction log analysis, final report to the Council on Library Resources: volume I of three volumes*. Dublin, Ohio: OCLC, 1983.
- United States. *Library of Congress Subject Headings*. 31 ed. Washington: Cataloging Distribution Service, Library of Congress, 2009.
- United States. *Encabezamientos de materia LC en español*. Washintong: Library of Congress, 2010. (DE, 25 de noviembre, 2010: <http://lcs-h-es.org>).
- Van Rijsbergen, et. al. "The selection of good search terms". *Information Processing and Management*. Oxford: Pergamon, vol. 17, núm.2, 1981. Pp. 77-91.
- Vega, Guadalupe y Víctor Cid. *Búsquedas fallidas en el uso del catálogo: cómo ayudar al usuario*. Documento presentado en las XXXVII Jornadas Mexicanas de Bibliotecología.
- Vickery, B.C. "Knowledge representation a brief review". En *Journal of Documentation*. Londres: Association of Special Libraries and Information Bureaux, vol. 42, núm. 3, 1986. Pp. 145-159.
- Villen Rueda, Luis. *La indización y el acceso por materias en los catálogos de bibliotecas: el desafío de la recuperación de información*. Jalisco, México: Universidad de Guadalajara, 2006.
- _____. "Indización y recuperación por materias en los OPACS de las bibliotecas españolas: ¿dos décadas de evaluación?". En *El profesional de la información*. Barcelona: Swets & Zeitlinger Ibérica, vol.15, núm. 2, 2006. Pp. 87-98.
- Wall, Eugene y Jean Barnes. *Intersystem compatibility and convertibility of subject vocabularies*. Philadelphia: Auerbach Corp, 1969.

Wersing, G. "The problematic situation as a basic concept of information science in the framework of the social sciences". En *New trends in informatics and its terminology*. Moscú: VINITI, 1979. Pp. 48-57.

Young, Heartsill. *Glosario ALA de bibliotecología y ciencias de la información*. Madrid: Diaz de Santos, 1988.

Ziman, John M. *Public knowledge: an essay concerning the social dimension of science*. Cambridge, Mass. : Universidad de Cambridge, 1968.

Anexos

Anexo 1. Transacciones de búsqueda por temas

Núm.	Petición de búsqueda
1	WSU= (afroamericanos and trabajos pictoricos)
2	WSU = (aulas)
3	WSU = (carceles)
4	WSU = (cartones)
5	WSU = (casa de moneda de chile)
6	WSU = (cristianos)
7	WSU = (musulmanes)
8	WSU = (diccionario and derecho)
9	WSU = (diseño organizacional)
10	WSU = (enciclopedias)
11	WSU = (estudios sobre cultura)
12	WSU = (Forget)
13	WSU = (Freshwater)
14	WSU = (history of brazil)
15	WSU = (IFE)
16	WSU = (impacto de la tecnologia)
17	WSU = (investigacion en periodismo)
18	WSU = (mecanismos de defensa)
19	WSU = (Participacion juvenil)
20	WSU = (perelman)
21	WSU = (policy networks)
22	WSU = (reclusion)
23	WSU = (reclusos)
24	WSU = (Soto y Gama)
25	WSU = (television corea)
26	WSU = (tuna)
27	WSU = (alcazar and segovia)
28	WSU = (archivo historico del agua)
29	WSU = (autismo)
30	WSU = (caribe and politica exterior)
31	WSU = (colorantes textiles)
32	WSU = (comunicacion estereotipos)
33	WSU = (corrido)
34	WSU = (cuerpo y comunicacion)
35	WSU = (danzas populares)
36	WSU = (evangelicos)
37	WSU = (HISTORIS)
38	WSU = (ingenios azucareros)
39	WSU = (INSTITUCIONES POLITICAS)

40	WSU = (leer)
41	WSU = (literatura mexicana)
42	WSU = (matute)
43	WSU = (mexico historia)
44	WSU = (pedofilia)
45	WSU = (periodica)
46	WSU = (place)
47	WSU = (seduccion)
48	WSU = (tanatologia)
49	WSU = (frases celebres and historia mexico)
50	WSU = (ascension)
51	WSU = (banamex)
52	WSU = (bolivia siglo xix)
53	WSU = (como sur historia)
54	WSU = (desarrollo de los niños)
55	WSU = (efren hernandez)
56	WSU = (estudiantes universitarios españa)
57	WSU = (instituciones filantropicas)
58	WSU = (HISPANA)
59	WSU = (historia porfiriato)
60	WSU = (iglesia evangelica en mexico)
61	WSU = (MARIPOSA)
62	WSU = (metacognicion)
63	WSU = (muerte edad media)
64	WSU = (porfiriato)
65	WSU = (Turistas)
66	WSU = (argumentacion)
67	WSU = (bolivia siglo xx)
68	WSU = (brass)
69	WSU = (desarrollo sostenible)
70	WSU = (enfermedades sexuales)
71	WSU = (oasis)
72	WSU = (prohibida)
73	WSU = (protestantes en mexico siglo xx)
74	WSU = (Conducta proxemica)
75	WSU = (pseudo)
76	WSU = (psicologia juridica)
77	WSU = (setting)
78	WSU = (televisiva)
79	WSU = (accesible)
80	WSU = (agenda)
81	WSU = (bacterias)
82	WSU = (cursos de estudio)

83	WSU = (cycle)
84	WSU = (desigualdad de genero)
85	WSU = (Hermenegildo Bustos)
86	WSU = (independencia)
87	WSU = (INDUSTRIA AUTOPARTES)
88	WSU = (limite)
89	WSU = (mexico y union europea)
90	WSU = (organicos)
91	WSU = (padres de familia)
92	WSU = (Transporte Sustentable)
93	WSU = (videojuegos)
94	WSU = (daniel cosio villegas)
95	WSU = (andres quintana roo)
96	WSU = (construction)
97	WSU = (masoquismo)
98	WSU = (comunicación and aspectos sociales and congresos)
99	WSU = (Belgium)
100	WSU = (boscan)
101	WSU = (caribe y mexico)
102	WSU = (DRAMATURGIA)
103	WSU = (flow)
104	WSU = (marquesado)
105	WSU = (mexico 1947)
106	WSU = (Pedro Castro)
107	WSU = (sociedad estudio de la)
108	WSU = (Sociologists)
109	WSU = (Wiener)
110	WSU = (Creditos)
111	WSU = (Criminal Law)
112	WSU = (Gonzalez Felipe)
113	WSU = (reliquias)
114	WSU = (rosales)
115	WSU = (SERVIDORES PUBLICOS)
116	WSU = (tematica)
117	WSU = (tourism)
118	WSU = (ultimas)
119	WSU = (goya)
120	WSU = (Liderazgo Mujeres)
121	WSU = (Obregon)
122	WSU = (seguridad ciudadana)
123	WSU = (migracion)
124	WSU = (olimpiadas)
125	WSU = (Venture)

126	WSU = (humanidades)
127	WSU = (INTERNET EN LA EDUCACION SUPERIOR)
128	WSU = (psicologia educativa)
129	WSU = (subempleo)
130	WSU = (controlados)
131	WSU = (economia de españa)
132	WSU = (fonseca)
133	WSU = (haciendas puebla)
134	WSU = (Landscape)
135	WSU = (motolinia)
136	WSU = (objeto)
137	WSU = (separacion iglesia estado)
138	WSU = (alcazar)
139	WSU = (curricula)
140	WSU = (INTERNET Y EDUCACION)
141	WSU = (METAMORFOSIS)
142	WSU = (niñez)
143	WSU = (Marine)
144	WSU = (virus)
145	WSU = (california gobierno historia)
146	WSU = (Cooperative)
147	WSU = (africa niños)
148	WSU = (dynamics)
149	WSU = (INDUSTRIA AUTOMOTRIZ MEXICO)
150	WSU = (loreto)
151	WSU = (perros)
152	WSU = (andalusia)
153	WSU = (machismo)
154	WSU = (Mothers)
155	WSU = (homosexual)
156	WSU = (padres de familia)
157	WSU = (truman)
158	WSU = (cossio)
159	WSU = (cultura y poder)
160	WSU = (Medicamentos)
161	WSU = (migrantes)
162	WSU = (perspectiva)
163	WSU = (tecnologias de la informacion)
164	WSU = (educacion basica)
165	WSU = (guerra de vietnam
166	WSU = (Analisis Estadistico)
167	WSU = (bolivia and siglo XIX and oruro)
168	WSU = (características nacionales cubanas)

169	WSU = (Alianzas estrategicas)
170	WSU = (elaboracion)
171	WSU = (ESTADO NACION)
172	WSU = (Garrido)
173	WSU = (iconografia and mexico)
174	WSU = (coloniales)
175	WSU = (Didactica)
176	WSU = (empresas electricas and mexico)
177	WSU = (marketing)
178	WSU = (societies)
179	WSU = (monitoreo)
180	WSU = (68)
181	WSU = (creatividad)
182	WSU = (discapacidad)
183	WSU = (personas con discapacidad)
184	WSU = (revoluciones and america latina)
185	WSU = (teatro griego)
186	WSU = (caribe and mexico)
187	WSU = (cronicas)
188	WSU = (nomenclatura)
189	WSU = (constructivismo)
190	WSU = (metodos de investigacion)
191	WSU = (reformas constitucionales)
192	WSU = (derecho aspectos politicos)
193	WSU = (Serrano)
194	WSU = (subjetividad)
195	WSU = (franquismo)
196	WSU = (Generacion)
197	WSU = (publicaciones)
198	WSU = (salud publica and congresos)
199	WSU = (VITICULTURA)
200	WSU = (asignacion temática)
201	WSU = (cientificos)
202	WSU = (BIOETICA)
203	WSU = (gay)
204	WSU = (mafia)
205	WSU = (politicas publicas)
206	WSU = (virtuales)
207	WSU = (Cholula)
208	WSU = (evangelizacion)
209	WSU = (MUERTOS E IGLESIA)
210	WSU = (catalogos)
211	WSU = (revolucion mexicana)

212	WSU = (Alianzas)
213	WSU = (leyes de reforma)
214	WSU = (Cadiz)
215	WSU = (numeros)
216	WSU = (conservation)
217	WSU = (iglesia evangelica)
218	WSU = (Tantrismo)
219	WSU = (Fusion)
220	WSU = (misterio y cuentos)
221	WSU = (cachemira)
222	WSU = (gonzalez y gonzalez, luis)
223	WSU = (infancia)
224	WSU = (diccionario)
225	WSU = (Diseño web)
226	WSU = (cuevas)
227	WSU = (Dolores)
228	WSU = (planificacion)
229	WSU = (empresas electricas)
230	WSU = (periodismo and mexico and historia)
231	WSU = (derechos del niño)
232	WSU = (mal)
233	WSU = (consolidacion)
234	WSU = (miedo)
235	WSU = (maximas and citas)
236	WSU = (coahuila historia)
237	WSU = (dia)
238	WSU = (space)
239	WSU = (africa and congo)
240	WSU = (santa anna)
241	WSU = (fundaciones)
242	WSU = (poder))
243	WSU = (españa medieval)
244	WSU = (paz mundial)
245	WSU = (imagenes)
246	WSU = (auditorias)
247	WSU = (negro)
248	WSU = (literatura colombiana)
249	WSU = (metafora)
250	WSU = (quimica)
251	WSU = (relacion)
252	WSU = (filosofia francesa)
253	WSU = (la biografia)
254	WSU = (medievales)

255	WSU = (lenguaje c)
256	WSU = (prisioneros)
257	WSU = (Germany)
258	WSU = (pueblos indigenas)
259	WSU = (tamaulipas historia)
260	WSU = (feminismo)
261	WSU = (libro)
262	WSU = (Jurisprudencia sociologica)
263	WSU = (trastornos)
264	WSU = (impacto)
265	WSU = (texas historia)
266	WSU = (expresion)
267	WSU = (fenomenologia)
268	WSU = (metodo)
269	WSU = (system)
270	WSU = (kant)
271	WSU = (nuevo leon historia)
272	WSU = (vietnam)
273	WSU = (anthropology)
274	WSU = (empresarial)
275	WSU = (relaciones publicas)
276	WSU = (patrimonio cultural)
277	WSU = (derecho)
278	WSU = (Hacienda)
279	WSU = (reproduccion)
280	WSU = (tradicion)
281	WSU = (Concepto)
282	WSU = (soberania)
283	WSU = (alvaro)
284	WSU = (Mercadotecnia)
285	WSU = (afroamericanos)
286	WSU = (sexualidad)
287	WSU = (niño)
288	WSU = (Desarrollo Sustentable)
289	WSU = (teoria de la informacion)
290	WSU = (politica exterior mexico)
291	WSU = (electronica)
292	WSU = (caricaturas)
293	WSU = (indigena)
294	WSU = (alfabetizacion)
295	WSU = (victor)
296	WSU = (Tomás)
297	WSU = (castas)

298	WSU = (chilena)
299	WSU = (presos)
300	WSU = (servicio civil)
301	WSU = (cuerpo)
302	WSU = (educativo)
303	WSU = (Daniel)
304	WSU = (lenguaje)
305	WSU = (geopolitica)
306	WSU = (romano)
307	WSU = (cine)
308	WSU = (zen)
309	WSU = (democracy)
310	WSU = (psicolinguistica)
311	WSU = (trabajos pictoricos)
312	WSU = (Villegas)
313	WSU = (Brasil -Siglo XIX)
314	WSU = (telecomunicaciones)
315	WSU = (santos)
316	WSU = (yucatan historia)
317	WSU = (italiana)
318	WSU = (historia de brasil)
319	WSU = (planeacion)
320	WSU = (literatura alemana)
321	WSU = (war)
322	WSU = (management)
323	WSU = (autonomia)
324	WSU = (personas)
325	WSU = (ciudadanía)
326	WSU = (aborto)
327	WSU = (fotografia)
328	WSU = (electoral)
329	WSU = (IMPUESTO)
330	WSU = (historiografia mexico)
331	WSU = (nutricion)
332	WSU = (colonialismo)
333	WSU = (espacio)
334	WSU = (coreana)
335	WSU = (Historia universal)
336	WSU = (andres)
337	WSU = (Constitucion)
338	WSU = (Turismo)
339	WSU = (muerte)
340	WSU = (amor)

341	WSU = (historia antigua)
342	WSU = (crimen)
343	WSU = (enfermedades)
344	WSU = (produccion)
345	WSU = (Felipe)
346	WSU = (mexico america latina)
347	WSU = (jesuitas)
348	WSU = (frontera)
349	WSU = (politica exterior)
350	WSU = (clasificacion)
351	WSU = (ELITE)
352	WSU = (Tlaxcala)
353	WSU = (organizaciones)
354	WSU = (political)
355	WSU = (genero)
356	WSU = (LEY)
357	WSU = (puebla historia)
358	WSU = (jean)
359	WSU = (william)
360	WSU = (desarrollo social)
361	WSU = (psicoanalisis)
362	WSU = (fernando)
363	WSU = (1806)
364	WSU = (prensa)
365	WSU = (periodismo)
366	WSU = (mexicanas)
367	WSU = (benito)
368	WSU = (revoluciones)
369	WSU = (calidad)
370	WSU = (instituciones)
371	WSU = (matematicas)
372	WSU = (juarez)
373	WSU = (redes)
374	WSU = (matrimonio)
375	WSU = (imperio)
376	WSU = (politicos)
377	WSU = (antropologia)
378	WSU = (conflicto)
379	WSU = (ingles diccionarios)
380	WSU = (urbanismo)
381	WSU = (opinion publica)
382	WSU = (mexico emigracion inmigracion)
383	WSU = (CHILE POLITICA GOBIERNO)

384	WSU = (valle)
385	WSU = (recursos naturales)
386	WSU = (CONQUISTA)
387	WSU = (Generales de guerra)
388	WSU = (interna)
389	WSU = (guerra civil)
390	WSU = (electronico)
391	WSU = (frances)
392	WSU = (1975)
393	WSU = (Globalizacion)
394	WSU = (RENTA)
395	WSU = (paz)
396	WSU = (mujer)
397	WSU = (demografia)
398	WSU = (politicas)
399	WSU = (budismo)
400	WSU = (arabe)
401	WSU = (politica ambiental)
402	WSU = (participacion ciudadana)
403	WSU = (arquitectura)
404	WSU = (los mexicanos)
405	WSU = (medio ambiente)
406	WSU = (migracion)
407	WSU = (historia argentina)
408	WSU = (estudios)
409	WSU = (nacionalismo)
410	WSU = (españoles)
411	WSU = (regionalismos)
412	WSU = (Bolivia)
413	WSU = (teatro)
414	WSU = (mapas)
415	WSU = (humanos)
416	WSU = (linguistica)
417	WSU = (Niños)
418	WSU = (empleo)
419	WSU = (familia)
420	WSU = (empresas)
421	WSU = (idioma)
422	WSU = (ingles)
423	WSU = (ciudades)
424	WSU = (central)
425	WSU = (pueblos)
426	WSU = (teoria)

427	WSU = (politica social)
428	WSU = (asia)
429	WSU = (tecnologia)
430	WSU = (recursos)
431	WSU = (gramatica)
432	WSU = (planificacion)
433	WSU = (popular)
434	WSU = (religion)
435	WSU = (derechos)
436	WSU = (africa)
437	WSU = (Español)
438	WSU = (ciencia)
439	WSU = (Enseñanza)
440	WSU = (Estadística)
441	WSU = (Argentina)
442	WSU = (Administracion)
443	WSU = (america latina)
444	WSU = (Brasil)
445	WSU = (conferencias)
446	WSU = (ensayos)
447	WSU = (economia)
448	WSU = (INDUSTRIA)
449	WSU = (desarrollo)
450	WSU = (filosofia)
451	WSU = (comercio)
452	WSU = (Estado)
453	WSU = (Congresos)
454	WSU = (educacion)
455	WSU = (LITERATURA)
456	WSU = (america)
457	WSU = (historia mexico)
458	WSU = (siglo)
459	WSU = (Politica)
460	WSU = (Mexico)
461	WSU = (Historia)
462	WSU = (cuidados previos al embarazo)
463	WSU = (Agüero and conflicto)
464	WSU = (aborto espontaneo)
465	WSU = (agenda setting)
466	WSU = (ainsa)
467	WSU = (alocuciones)
468	WSU = (alocuciones ensayos y conferencias)
469	WSU = (alvaro matute el 68 y la historia)

470	WSU = (alvaro matute estudios historiograficos)
471	WSU = (ANIMALES- METAMORFOSIS)
472	WSU = (antagonistas)
473	WSU = (anthropology of space and place)
474	WSU = (antropologia-evolucionismo)
475	WSU = (asignacion tematica)
476	WSU = (aulas virtuales)
477	WSU = (autonomia indigena en Bolivia)
478	WSU = (bacterias antagonistas)
479	WSU = (baudelaire y el artista de la vida)
480	WSU = (baudelaire y el artista de la vida and azua)
481	WSU = (biodiversity)
482	WSU = (bonnin)
483	WSU = (calidad del empleo en mexico)
484	WSU = (canada political system)
485	WSU = (caso de recursos humanos)
486	WSU = (catalogo banamex)
487	WSU = (centrifugacion)
488	WSU = (chales)
489	WSU = (chido)
490	WSU = (Cholula urbanismo ciudades colonia)
491	WSU = (Chuzo)
492	WSU = (ciudades coloniales tlaxcala)
493	WSU = (clasificacion de empresas)
494	WSU = (clasificacion de empresas segun la forma de explotacion)
495	WSU = (clonazepam)
496	WSU = (cognositivo)
497	WSU = (comunicologia)
498	WSU = (Concepto Alianzas Estrategicas)
499	WSU = (conflicto ambiental)
500	WSU = (conquista arabe)
501	WSU = (consecuencias)
502	WSU = (Constitucion and Cadiz and Mexico)
503	WSU = (Contribucion a la critica de la economia politica)
504	WSU = (Cooperative societies management)
505	WSU = (Creditos administracion)
506	WSU = (crimen y mafia italiana)
507	WSU = (cronicas medievales)
508	WSU = (cuerpo y expresion)
509	WSU = (culinaria)
510	WSU = (democracy in the caribbean)
511	WSU = (derechos de los pueblos indigenas)
512	WSU = (desarrollo cognositivo del niño)

513	WSU = (desarrollo regional en mexico)
514	WSU = (dia de la ascension)
515	WSU = (Didactica general)
516	WSU = (discretas)
517	WSU = (disoluciones)
518	WSU = (DIVINOS)
519	WSU = (DRAMATURGIA DE MUERTOS)
520	WSU = (ecosystem)
521	WSU = (educacion de prisioneros)
522	WSU = (elaboracion de cursos)
523	WSU = (el concepto de lo politico)
524	WSU = (el imperio romano antes de la era cristiana)
525	WSU = (El metodo cientifico and 2006)
526	WSU = (el sacerdote misterio de amor)
527	WSU = (El turismo accesible)
528	WSU = (El turismo accesible como una forma de integracion a la vida social de las personas)
529	WSU = (el ultimo de los mohicanos)
530	WSU = (empleadores)
531	WSU = (empleadores de personas con discapacidad)
532	WSU = (enciclopedia mecatronica)
533	WSU = (enciclopedias de marketing)
534	WSU = (enfermedades sexuales en mexico colonial)
535	WSU = (Enseñanza universitaria-Recursos en redes de computacion)
536	WSU = (espacio para la infancia)
537	WSU = (Español niños migrantes)
538	WSU = (espontaneo)
539	WSU = (Estadística Intermedia)
540	WSU = (estadística y su objeto de estudio)
541	WSU = (estados del sur e u a)
542	WSU = (estudios sobre la familia mexicana)
543	WSU = (evangelicos en mexico)
544	WSU = (evolucionismo)
545	WSU = (familia limitrofe)
546	WSU = (familia mexicana)
547	WSU = (ficha electronica)
548	WSU = (frases celebres)
549	WSU = (Freshwater biodiversity conservation)
550	WSU = (fundaciones civicas)
551	WSU = (Fusion y consolidacion)
552	WSU = (gay tourism)
553	WSU = (Generacion x chilena)
554	WSU = (genero y ciudadania)

555	WSU = (Germany second world war)
556	WSU = (Hacienda Dolores)
557	WSU = (hepatistis)
558	WSU = (hepatistis A virus)
559	WSU = (historia del evolucionismo)
560	WSU = (historiograficos)
561	WSU = (HISTORIS DE LA CONQUISTA)
562	WSU = (homogenizacion)
563	WSU = (homogenizacion y centrifugacion)
564	WSU = (hopital)
565	WSU = (hopital de san andres)
566	WSU = (hospocios)
567	WSU = (IETU)
568	WSU = (impacto del comercio intraindustrial en la industria manufacturera)
569	WSU = (impuesto empresarial a tasa unica)
570	WSU = (IMPUESTO SOBRE LA RENTA 1975)
571	WSU = (interespecifica)
572	WSU = (Ion)
573	WSU = (Ion flow dynamics)
574	WSU = (jesuitas siglo XVI-XVII)
575	WSU = (JUICIOS DIVINOS)
576	WSU = (La araucana)
577	WSU = (la fenomenologia de la religion)
578	WSU = (La literatura chilena actual)
579	WSU = (la mujer en el porfiriato)
580	WSU = (Landscape construction)
581	WSU = (la participacion politica de la mujer en el porfiriato)
582	WSU = (la revolucion en el estado de nuevo leon)
583	WSU = (leer es chido)
584	WSU = (lenguajes controlados)
585	WSU = (libro electronico)
586	WSU = (LITERATURA HISPANA)
587	WSU = (Long Swings and the nonproductive cycle)
588	WSU = (los jesuitas en mexico)
589	WSU = (ludoteca)
590	WSU = (mal frances)
591	WSU = (mapas de asia central)
592	WSU = (Marine biodiversity conservation)
593	WSU = (Marine ecosystem management)
594	WSU = (MARIPOSA- METAMORFOSIS)
595	WSU = (matematicas discretas)
596	WSU = (Mcondo)
597	WSU = (mecatronica)

598	WSU = (Medicamentos administracion)
599	WSU = (medio ambiente e indigenas)
600	WSU = (meixoco)
601	WSU = (metodo de cross)
602	WSU = (metodos indirectos)
603	WSU = (mexico -- geopolitica)
604	WSU = (monitoreo de instituciones)
605	WSU = (morfologia)
606	WSU = (narco)
607	WSU = (niñez e independencia)
608	WSU = (niños en la independencia)
609	WSU = (nonproductive)
610	WSU = (numeros primos)
611	WSU = (nutricion curricular)
612	WSU = (ordalias)
613	WSU = (organizaciones filantropicas)
614	WSU = (participacion electoral en las ultimas elecciones)
615	WSU = (peninsula coreana)
616	WSU = (perelmtoulmin)
617	WSU = (perro)
618	WSU = (perspectiva de genero)
619	WSU = (planificacion de la auditoria interna)
620	WSU = (politica borbonica en Chile)
621	WSU = (politica en el idioma ingles)
622	WSU = (politicas de telecomunicaciones)
623	WSU = (postulados)
624	WSU = (postulados generales de la ciencia)
625	WSU = (produccion televisiva)
626	WSU = (pseudociencia)
627	WSU = (reclusorios)
628	WSU = (recubrimientos)
629	WSU = (recubrimientos organicos)
630	WSU = (recursos municipios)
631	WSU = (relacion interespecifica de lacertilios)
632	WSU = (religion prensa educacion)
633	WSU = (reproduccion prohibida)
634	WSU = (Rubem)
635	WSU = (Rubem fonseca)
636	WSU = (salmonelosis)
637	WSU = (santos e imagenes)
638	WSU = (santos y evangelizacion)
639	WSU = (ser and subjetividad)
640	WSU = (Should Sociologists Forget Their Mothers and Fathers)

641	WSU = (sistema educativo mexicano)
642	WSU = (subjetividad and kant))
643	WSU = (sustentabilidad)
644	WSU = (Swings)
645	WSU = (tecnologia y sus consecuencias)
646	WSU = (telenovela)
647	WSU = (telenovela coreana)
648	WSU = (teoria de la argumentacion)
649	WSU = (testro)
650	WSU = (testro popular america latina)
651	WSU = (Tlaxcala arquitectura ciudad)
652	WSU = (tlcan)
653	WSU = (tradicion culinaria)
654	WSU = (trastornos limite)
655	WSU = (turismo accesible)
656	WSU = (UNIFEM)
657	WSU = (urbanismo colonial tlaxcala)
658	WSU = (vouyerismo)
659	WSU = (zapote)
660	WSU = (zapote negro)

Anexo 2: Cuestionario

CUESTIONARIO

Objetivo:

Indagar sobre las implicaciones del encuentro entre el lenguaje del usuario y el lenguaje de indización en la recuperación temática en el catálogo público de la BDCV.

Instrucciones:

I. Ordene las opciones que se le presentan. Asigne el número 1 a la opción más frecuente y los números consecutivos a las opciones siguientes.

1. En el catalogo de la biblioteca, la mayoría de las veces, busco mediante:

- Búsqueda básica
- Búsqueda avanzada
- Búsqueda por índices
- Búsqueda limitada
- Búsqueda por comandos

2. De las opciones que presenta el catálogo, la mayoría de las veces, busco por:

- Todos los campos
- Título
- Autor
- Temas
- ISBN
- Editorial
- Notas
- Serie

II. En las preguntas cerradas subraye la opción que más se adapte a su experiencia usando el catálogo de la biblioteca. Mientras que en las preguntas abiertas, argumente detalladamente su respuesta.

3. La búsqueda por temas la uso:

- a) Siempre
- b) Casi siempre
- c) Algunas veces
- d) Pocas veces
- e) Nunca

¿Por qué?

4. Buscar por temas me resulta:

- a) Muy fácil

- b) Fácil
- c) Difícil
- d) Muy difícil

¿Por qué?

5. Si comparamos la búsqueda por temas con la búsqueda por título, la búsqueda por temas es:

- a) Más fácil
- b) Más difícil

¿Por qué?

6. Si comparamos la búsqueda por temas con la búsqueda por autor, la búsqueda por temas es:

- a) Más fácil
- b) Más difícil

¿Por qué?

7. Lo más complicado en la búsqueda por temas es:

- a) Encontrar la opción de búsqueda por temas en el catálogo
- b) Elegir la palabra o palabras con las que iniciaré la búsqueda
- c) Recuperar documentos a partir de la palabra con la que busqué
- d) Seleccionar de los registros, que arroja el catálogo, los más útiles

8. Concretar en palabras o frases el tema que estoy buscando es:

- a) Muy fácil
- b) Fácil
- c) Difícil
- d) Muy difícil

¿Por qué?

9. Antes de empezar la búsqueda por temas, la mayoría de las veces:

- a) No tengo ni idea de qué palabra o palabras usar para empezar la búsqueda
- b) Tengo una palabra que representa de manera general lo que necesito
- c) Tengo una palabra que representa con exactitud lo que necesito
- d) Tengo un descriptor que representa lo que necesito
- e) Tengo un encabezamiento de materia que representa lo que necesito

10. Encontrar la palabra o palabras con las que iniciaré la búsqueda por temas me resulta:

- a) Sencillo y siempre encuentro la adecuada
- b) Sencillo, pero no siempre es la adecuada
- c) Complicado, pero siempre encuentro la adecuada
- d) Complicado y no siempre es la adecuada

11. ¿En qué parte del registro completo, que muestra el catalogo, aparecen los temas que la biblioteca asigna a los documentos?

12. Cuando consulto el registro, reviso estos temas que asigna la biblioteca:

Sí ____ No__

*En caso afirmativo, con qué frecuencia:

- a) Siempre
- a) Casi siempre
- b) Algunas veces
- c) Pocas veces

13. Los temas asignados por la biblioteca, generalmente, coinciden con las palabras que uso para buscar:

- a) Siempre
- b) Casi siempre
- c) Algunas veces
- d) Pocas veces
- e) Nunca
- f) No sé, porque no los reviso

14. Los términos que uso para buscar por temas generalmente son:

- a) Una palabra clave
- b) Dos o más palabras enlazadas
- c) Una frase sobre el tema
- d) Un descriptor
- e) Un encabezamiento temático

15. A partir de las palabras que uso para buscar por temas, recupero documentos útiles:

- a) Siempre
- b) Casi siempre
- c) Algunas veces
- d) Pocas veces
- e) Nunca

16. De los documentos que recupero con la palabra de búsqueda seleccionada, generalmente:

- a) Todos tratan el tema que estaba buscando
- b) En su mayoría tratan el tema que estaba buscando
- c) Pocos tratan el tema que estaba buscando
- d) Ninguno trata el tema que estaba buscando

17. Cuando mi búsqueda por temas no da resultados, pienso que:

- a) Hice algo mal en el procedimiento de búsqueda
- b) La biblioteca no tiene documentos sobre el tema
- c) La palabra con la que busqué no fue la adecuada
- d) Algo está funcionando mal en el catálogo
- e) Otra, ¿cuál? _____

18. Cuando mi búsqueda por temas no da resultados:

- a) Abandono la búsqueda
- b) Busco con otras palabras (sinónimos) del tema
- c) Busco temas relacionados
- d) Consulto vocabularios controlados (listas de encabezamientos, tesauros, etc.)
- e) Solicito la ayuda de un bibliotecario

19. Siento que tengo problemas para recuperar documentos, cuando busco:

- a) Temas generales de mi área de estudio o de investigación
- b) Temas específicos de mi área de estudio o investigación
- c) Temas nuevos de mi área de estudio o investigación
- d) Nunca tengo problemas para encontrar temas de mi área de estudio

20. ¿Actualmente está realizando una investigación o trabajo académico para el cual necesite usar el catálogo de la biblioteca?

Sí _____

No _____

En caso afirmativo, describa brevemente el tema de su trabajo o investigación:

21. ¿Con qué palabras buscaría ese tema en el catálogo de la biblioteca?

- 1 _____
- 2 _____
- 3 _____
- 4 _____
- 5 _____

22. Según su opinión, ¿qué debe hacer la biblioteca para que los usuarios tengan mejores resultados en las búsquedas por temas?

¡Gracias por su colaboración!

Comentarios o sugerencias al cuestionario:
