



FACULTAD DE LETRAS Y CIENCIAS HUMANAS

**DETERMINACIÓN DE LAS CAUSAS DE BÚSQUEDAS SIN RESULTADOS
EN UN CATÁLOGO BIBLIOGRÁFICO EN LÍNEA MEDIANTE EL ANÁLISIS
DE BITÁCORAS DE TRANSACCIONES: EL CASO DE LA PONTIFICIA
UNIVERSIDAD CATÓLICA DEL PERÚ**

Tesis para optar el título de Licenciado en Bibliotecología y Ciencias de la Información
que presenta el Bachiller:

Raúl Alejandro Sifuentes Arroyo

Asesora: Mónica Arakaki Heshiki

Lima, Perú - 2013

RESUMEN

Determinación de las causas de búsquedas sin resultados en un catálogo bibliográfico en línea mediante el análisis de bitácoras de transacciones: el caso de la Pontificia Universidad Católica del Perú

El presente trabajo de investigación está orientado a determinar las causas de las búsquedas sin resultados realizadas durante todo el año 2012 en el catálogo en línea del Sistema de Bibliotecas de la Pontificia Universidad Católica del Perú. Para este fin se hizo uso del análisis de bitácora de transacciones de búsqueda como metodología, tras el cual se llegó a identificar tres causas que generaron recuperación vacía en las búsquedas: 1) Términos no coincidentes: palabras o frases escritas correctamente en las fórmulas de búsqueda que no coinciden con las usadas en la descripción bibliográfica y terminología temática asignada a cada registro. 2) Escritura incorrecta: referida a los errores tipográficos y ortográficos en las fórmulas de búsqueda. 3) Elección del índice incorrecto: cuando el usuario selecciona un índice no pertinente para su fórmula de búsqueda. Se hizo un análisis más detallado para cada uno de los índices de búsqueda que ofrece el catálogo en línea. A la luz de los resultados, se sugieren aspectos a reforzar en las capacitaciones de uso del catálogo en línea, mejoras en la interfaz de búsqueda y uso de las bitácoras de transacciones de búsqueda para detectar demanda de recursos bibliográficos que las colecciones no satisfacen.

Palabras clave: Análisis de bitácoras de transacciones, Estudios de usuarios, Búsquedas bibliográficas, Recuperación de la información, LTA, Catálogos en línea, OPAC.

TABLA DE CONTENIDOS

INTRODUCCIÓN.....	1
CAPÍTULO 1: MARCO TEÓRICO	6
1.1. Los estudios de conducta de información	6
1.2. La recuperación de información.....	7
1.3. Los catálogos en línea	9
1.4. El análisis de las bitácoras de transacciones	10
1.4.1. Definición	10
1.4.2. Dificultades de su uso	11
1.4.3. Tipos de análisis	12
1.4.4. Estudios de uso de catálogos en línea mediante el análisis de bitácoras de transacciones	14
CAPÍTULO 2: EL ENTORNO.....	21
2.1. El Sistema de Bibliotecas de la PUCP.....	21
2.1.1. Servicios brindados.....	21
2.1.2. La colección bibliográfica	22
2.1.3. Los usuarios	23
2.2 El catálogo en línea	24
2.2.1 Características y funciones	24
2.2.2 El motor de búsqueda	26
CAPITULO 3: METODOLOGÍA DE ANÁLISIS.....	29
3.1. Acerca de la metodología	29
3.2. Las bitácoras de transacciones de búsqueda de e-Library®	30
3.2.1. Descripción de los elementos	30
3.2.2. Muestra de datos originales	32
3.2.3. Muestra de datos procesados	33

3.3. Recolección de los datos	34
3.4. Herramientas de análisis.....	35
CAPITULO 4: RESULTADOS DEL ANÁLISIS.....	37
4.1. Número total de búsquedas.....	37
4.2. Proporción de búsquedas con resultados vacíos	39
4.3. Número de búsquedas por tipo (simple, avanzada)	41
4.4. Búsquedas por índice (general, autor, título, tema, revista-periódico, serie, editorial)	42
4.5. Proporción de resultados vacíos en las búsquedas por el número de palabras usadas (longitud de las consultas).....	45
4.6. Identificación de causas que generaron recuperación vacía en las búsquedas	47
4.6.1 Causas de búsquedas sin resultados según índices usados durante las semanas con mayor número de búsquedas	49
4.6.1.1. Causas de búsquedas sin resultados por uso del índice <i>tema</i>	49
4.6.1.2. Causas de recuperación vacía para las búsquedas en el índice <i>título</i>	51
4.6.1.3. Causas de recuperación vacía para las búsquedas por combinación de índices.....	51
4.6.1.4. Causas de recuperación vacía para las búsquedas en el índice <i>general</i> ..	52
4.6.1.5 Causas de recuperación vacía para las búsquedas en el índice de <i>autor</i> ..	52
4.6.1.6 Causas de recuperación vacía para las búsquedas en el índice de <i>revista-periódico</i>	53
4.6.1.7 Causas de recuperación vacía para las búsquedas en el índice de <i>serie</i> ...	53
CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES	56
5.1. Conclusiones	56
5.1.1. Sobre el análisis de bitácoras de transacciones como metodología	56
5.1.2. Sobre la interacción del usuario y el catálogo en línea del Sistema de Bibliotecas de la PUCP	57
5.1.3. Sobre las búsquedas sin resultados	58
5.2 Recomendaciones	60

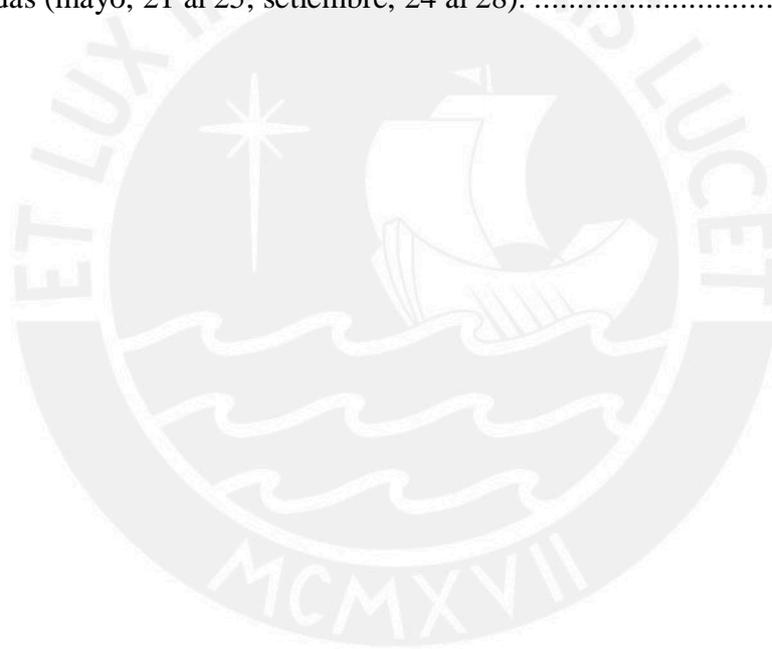
5.2.1. Capacitaciones de uso del catálogo en línea	60
5.2.2. Mejoras en la interfaz de búsqueda	60
5.2.3. Activar el módulo de control de autoridades temáticas en el catálogo	61
5.2.4. Estudios mensuales de términos no coincidentes para mejorar el desarrollo de colecciones	62
5.2.5. Estudios complementarios a la presente investigación	62
5.3 Comentario final.....	63
BIBLIOGRAFÍA	i
ANEXO 1: Proporción mensual de búsquedas con resultados vacíos para el año 2012.. v	
ANEXO 2: Detalle de la proporción mensual de búsquedas con resultados vacíos.....	vi
ANEXO 3: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>general</i> (2012)	vii
ANEXO 4: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>autor</i> (2012)	viii
ANEXO 5: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>título</i>	ix
ANEXO 6: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>tema</i> (2012)	x
ANEXO 7: Detalle mensual de la proporción de resultados vacíos en las búsquedas de índices combinados	xi
ANEXO 8: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>revista- periódico</i> (2012).....	xii
ANEXO 9: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>serie</i> (2012)	xiii
ANEXO 10: Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice <i>editorial</i> (2012).....	xiv
ANEXO 11: Detalle mensual de la proporción de resultados vacíos en las búsquedas según el número de palabras usadas (2012)	xv
ANEXO 12: Comparativo del número de búsquedas realizadas en el catálogo en línea de la PUCP para los años 2009 y 2010	xvi

ANEXO 13: Comparativo del número de búsquedas realizadas en el catálogo en línea de la PUCP para los años 2010 y 2011	xvii
ANEXO 14: Comparativo del número de búsquedas realizadas en el catálogo en línea de la PUCP para los años 2011 y 2012	xviii
ANEXO 15: <i>Script</i> que da formato a las bitácoras de transacciones de búsqueda del sistema Unicorn o Symphony	xix



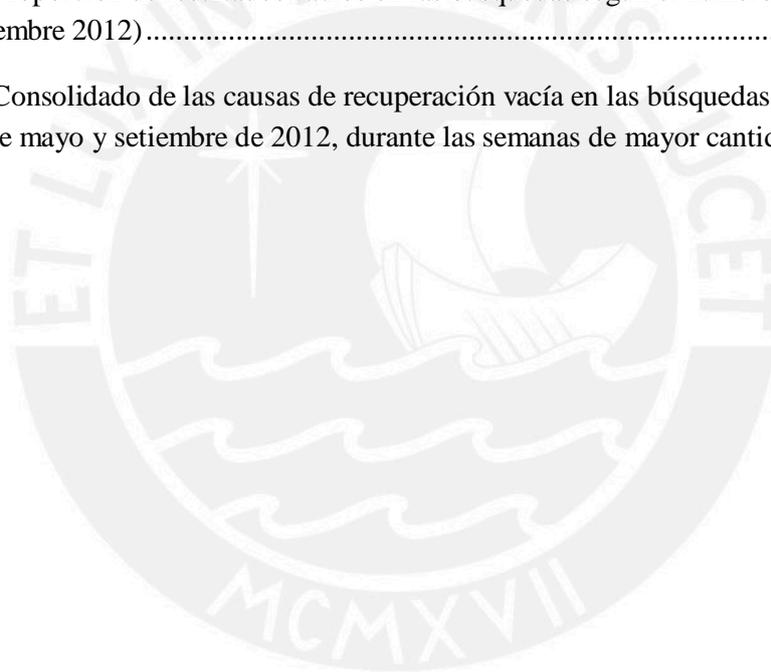
Lista de tablas

Tabla 1: Número de búsquedas totales para el año 2012	39
Tabla 2: Proporción de resultados vacíos en búsquedas por palabra clave el año 2012 .	39
Tabla 3: Número de búsquedas por tipo	41
Tabla 4: Proporción de resultados vacíos de las búsquedas en todos los índices en relación a la frecuencia de su uso para el periodo de enero a diciembre del 2012.	43
Tabla 5: Proporción de resultados vacíos en las búsquedas según el número de palabras usadas (enero-diciembre 2012)	46
Tabla 6: Consolidado de las causas de recuperación vacía en las búsquedas por índices usados los meses de mayo y setiembre de 2012, durante las semanas de mayor cantidad de búsquedas (mayo, 21 al 25; setiembre, 24 al 28).	54



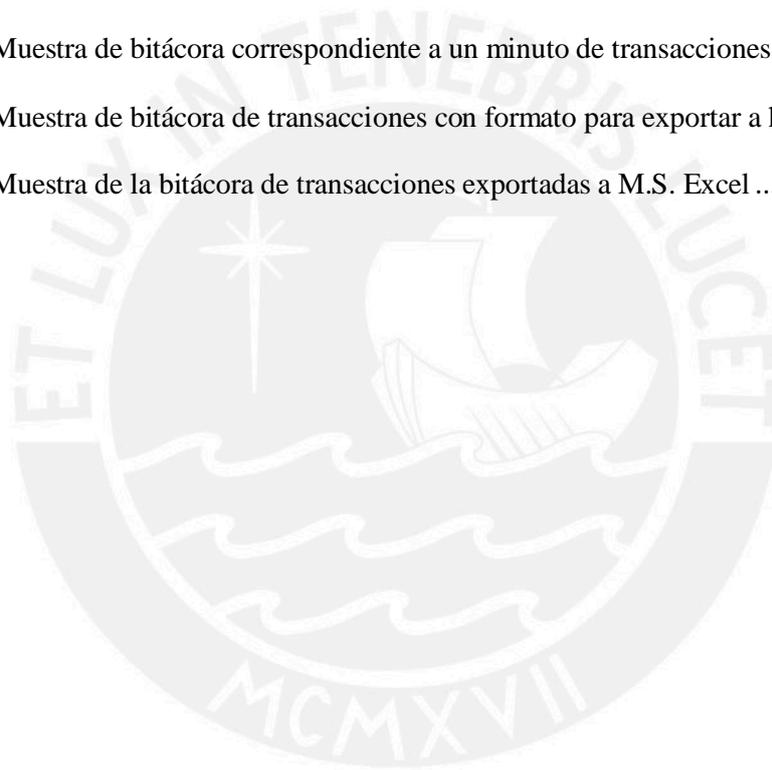
Lista de gráficos

Gráfico 1: Comparativo mensual de número de búsquedas para los años 2011 y 2012.....	38
Gráfico 2: Comparativo de sesiones por horas del día en e-Library para los años 2011 y 2012.	38
Gráfico 3: Comparación mensual del número de búsquedas frente al número de búsquedas con resultados vacíos	40
Gráfico 4: Proporción mensual de búsquedas sin resultados	40
Gráfico 5: Proporción de búsquedas con resultados vacíos por tipo de búsqueda	41
Gráfico 6: Relación entre la frecuencia de uso de índices y el porcentaje de recuperación vacía para el periodo de enero a diciembre de 2012.	45
Gráfico 7: Proporción de resultados vacíos en las búsquedas según el número de palabras usadas (enero-diciembre 2012).....	47
Gráfico 8: Consolidado de las causas de recuperación vacía en las búsquedas por índices usados los meses de mayo y setiembre de 2012, durante las semanas de mayor cantidad de búsquedas	55



Lista de imágenes

Imagen 1: Pantalla del formulario de búsqueda simple	27
Imagen 2: Pantalla del formulario de búsqueda avanzada	28
Imagen 3: Pantalla de resultados de búsqueda con opción "Que empiece con" (Búsqueda alfabética).....	28
Imagen 4: Muestra de bitácora de transacciones de inicio de sesión	31
Imagen 5: Muestra de bitácora de transacciones con información adicional (caso 1).....	32
Imagen 6: Muestra de bitácora de transacciones con información adicional (caso 2).....	32
Imagen 7: Muestra de bitácora correspondiente a un minuto de transacciones	33
Imagen 8: Muestra de bitácora de transacciones con formato para exportar a hoja de cálculo...34	
Imagen 9: Muestra de la bitácora de transacciones exportadas a M.S. Excel	34



INTRODUCCIÓN

En 1997 entró en servicio el catálogo en línea del Sistema de Bibliotecas de la Pontificia Universidad Católica del Perú, en adelante PUCP, a través de la interfaz llamada Webcat®, constituyéndose en el primer catálogo en línea del Perú basado en Web. Nueve años después es reemplazada por iLink®, novedosa interfaz que permitía complementar automáticamente la información de las fichas bibliográficas con tablas de contenido, resúmenes y cubiertas de libros. En el 2009, esta última se reemplaza por e-Library®, interfaz mejorada en estilo y nuevas funciones, entre las que destaca la adecuación a dispositivos móviles.

Desde el año 2002, la Unidad de Automatización del Sistema de Bibliotecas de la PUCP ha realizado estudios de uso del catálogo en línea mediante análisis de las bitácoras del servidor Web, donde se registran importantes datos con los cuales se han obtenido cifras anuales sobre el número total de búsquedas, el número de sesiones, los tipos de búsquedas, el número de búsquedas por horas, entre otros. Llamó mucho la atención desde entonces el elevado número de búsquedas realizado por los usuarios, el cual llegó a pasar los 3.8 millones (Sistema de Bibliotecas de la Pontificia Universidad Católica del Perú. Sistema de Bibliotecas 2007: 44), con lo que el catálogo demostró ser la herramienta de búsqueda bibliográfica más usada.

Con la instalación de iLink® fue posible registrar las transacciones de búsquedas en bitácoras de texto. Al cabo de aproximadamente un año de funcionamiento se recolectó y analizó una muestra de esas bitácoras cuyos resultados causaron preocupación ya que poco más de un tercio de las búsquedas realizadas en el catálogo no obtuvo resultados. Se plantearon entonces las siguientes preguntas:

- ¿Se realizan muchas búsquedas porque en primera instancia los usuarios no encuentran lo que necesitan?
- ¿Qué porcentaje de las búsquedas totales no obtiene resultados?
- ¿Cuál es el tipo de búsqueda que registra más resultados vacíos?
- ¿Qué índice registra el desempeño más bajo según el número de resultados vacíos que obtiene?
- ¿Cuáles son las causas de las búsquedas sin resultados?

La elevada proporción de búsquedas sin resultados justifica la necesidad de llevar cabo un estudio más detallado a fin de determinar su relación con la manera cómo buscan los usuarios del catálogo en línea de las bibliotecas de la PUCP. La presente investigación se justifica, además, porque va a permitir proponer mejoras en la interfaz de búsquedas, así como también en las capacitaciones que brinda a sus usuarios el Sistema de Bibliotecas de la PUCP, y porque es un aporte para la bibliotecología en el país ya que no se ha encontrado estudios publicados donde se haya usado el análisis de las bitácoras de transacciones de los catálogos en línea como metodología.

La hipótesis de la presente investigación es la siguiente: las causas de las búsquedas sin resultados están relacionadas en gran medida con errores que cometen los usuarios durante sus búsquedas.

El análisis de las bitácoras de transacciones de búsqueda va a permitir determinar las causas que llevan a obtener resultados vacíos como respuesta a las consultas hechas por los usuarios del catálogo. Ese es el primer objetivo del estudio.

El segundo objetivo es la elaboración de un conjunto de propuestas de mejoras para el catálogo en línea y la interfaz de búsqueda, para reforzar aspectos en las capacitaciones que se brindan a los alumnos ingresantes y a otros alumnos que las solicitan, así como también para la elaboración de ayudas en línea y/o impresas.

Para el presente trabajo de investigación se ha recolectado y analizado las bitácoras de las transacciones de búsqueda del catálogo en línea del Sistema de Bibliotecas de la PUCP, específicamente aquellas generadas durante todo el año 2012, para todas las bibliotecas, incluyendo la de CENTRUM Católica, la de Instituto Riva-Agüero y la del Instituto Confucio. En total fueron alrededor de cuatro millones y medio de líneas de texto, donde se encontraron 2'866,649 de transacciones de búsquedas, las cuales fueron analizadas durante aproximadamente 3 meses desde enero del 2013. Se hizo el mismo ejercicio de análisis para cada uno de los meses del año 2012 a fin de establecer patrones de comportamiento en las búsquedas de los usuarios del catálogo en línea, específicamente en aquellas donde no obtenían resultados. No se ha podido realizar un estudio más específico, delimitado, por ejemplo, a ciertos perfiles de usuarios debido a que la gran mayoría de transacciones de búsquedas se hacen dentro de sesiones anónimas, donde el usuario no se identifica en sus sesiones de búsqueda porque el Sistema de Bibliotecas de la PUCP no tiene por política la identificación obligatoria del usuario para este servicio. Sin embargo, con fines de estudios futuros se podría obligar a los usuarios a identificarse en sus sesiones de búsqueda y de esta manera poder tener bitácoras que incluyan el identificador del usuario, dato valioso que permitiría cruzar información con la base de datos de usuarios y enriquecer los datos que proporcionan las bitácoras.

Para poder analizar las bitácoras es necesario darles un formato adecuado para poder procesar los datos en hojas de cálculo. Esto puede ser una dificultad para quienes no están familiarizados en el procesamiento de datos, especialmente en entornos

Linux/Unix. Para la presente investigación se tuvo la suerte de encontrar un *script*¹ hecho en lenguaje de programación PERL, disponible en el repositorio de la comunidad de desarrolladores de la empresa SirsiDynix, con el cual se pudo dar formato a los datos para su fácil exportación y procesamiento en M.S. Excel.

El primer capítulo está dedicado al marco teórico que sustenta y guía todo el trabajo de investigación realizado. Se parte de los llamados estudios de conducta de información, donde se enfoca el tema de la recuperación de información en una perspectiva histórica, clave para entender la importancia del presente estudio. De igual manera se presenta una breve historia de los catálogos bibliográficos en línea, señalando los métodos de recuperación de información que estos fueron incorporando. Lo anterior para poder comprender la naturaleza de una bitácora de transacciones de búsquedas y dar paso a su definición, las dificultades de su uso y los estudios realizados en otras instituciones, muy útiles para la selección de algunos modelos de análisis que se han aplicado al presente trabajo.

En el segundo capítulo se brinda información sobre el entorno al cual pertenece la interfaz de búsqueda objeto de este estudio. Se describe brevemente cómo está organizado el Sistema de Bibliotecas de la PUCP, los servicios que brinda, su colección y sus usuarios. Con detalle se describen las principales características y funciones del catálogo en línea y su motor de búsquedas.

En el tercer capítulo se describe la metodología usada en el presente trabajo de investigación y se detalla cada uno de los elementos que conforman las bitácoras de transacciones de búsquedas del catálogo en línea del Sistema de Bibliotecas de la PUCP, se incluye una muestra original de los datos y otra muestra de los mismos pero

¹ Un script es un archivo de texto que contiene una serie de instrucciones que ejecutan una tarea o función. La estructura y sintaxis dependen del lenguaje de programación en el que está escrito.

ya procesados para el análisis en hojas de cálculo. Se explica el proceso de recolección de los datos y las herramientas que se usaron para el análisis.

En el cuarto capítulo se presentan los resultados del análisis, desde lo más general a lo más específico, donde se determinan las causas de las búsquedas sin resultados para cada uno de los índices de búsqueda analizados.

En el quinto y último capítulo se presentan las conclusiones sobre el análisis de bitácoras de transacciones como metodología, sobre la interacción del usuario y el catálogo en el Sistema de Bibliotecas de la PUCP, así como también sobre las búsquedas sin resultados. Se concluye con las recomendaciones orientadas a reducir la proporción de búsquedas sin resultados, así como también propuestas de uso del análisis de bitácoras de transacciones para apoyar el trabajo de desarrollo de colecciones, y para la realización de otros estudios, siempre orientados a mejorar los servicios de búsqueda de las bibliotecas.

CAPÍTULO 1: MARCO TEÓRICO

1.1. Los estudios de conducta de información

La recuperación de información ha sido estudiada ampliamente, incluso antes de la aparición de Internet, dentro de lo que se conoce como '*conducta de información*' (*information behaviour*), término acuñado por el británico Thomas D. Wilson para referirse a las actividades en las que un usuario de información se involucra cuando determina sus necesidades de información, la busca de distintas maneras, la usa o la transfiere (1999: 249). Presenta modelos de investigación en '*conducta de información*' y hace un repaso de los distintos modelos existentes, hasta ese momento, para lo que se conoce como '*conducta de requerimiento de información*' (*information-seeking behaviour*) y '*conducta en búsqueda de información*' (*information search behaviour*). Sugiere un modelo anidado donde '*conducta en búsqueda de información*' forma parte de '*conducta de requerimiento de información*', y este último, de '*conducta de información*' (Wilson 1999: 262-263)². La '*conducta de requerimiento de información*' es un conjunto de actividades que se inician con una necesidad determinada por un usuario, quien, para satisfacerla, acude a fuentes de información formales o informales, o servicios, con resultados exitosos o fallidos, con la finalidad de encontrar información relevante (Wilson 1999: 251). La '*conducta de búsqueda de información*' particularmente se refiere a las interacciones entre el usuario de información (con o sin intermediario) y sistemas de información basados en computadoras, de los cuales los sistemas de recuperación de información de datos textuales pueden ser considerados como un tipo. (Wilson 1999: 263).

² La traducción al español de las palabras inglesas *search* y *seeking* podrían causar confusión ya que ambas suelen traducirse de manera similar. En función de las definiciones dadas por el autor se ha optado por darle el significado de '*requerir*' al verbo inglés *seek*.

1.2. La recuperación de información

En un reciente artículo, Sanderson y Croft compilan de forma breve la historia de la recuperación de información e incluso llegan a especular sobre las futuras direcciones que tomará este campo y subrayan que la historia de la recuperación de información no empieza con la Internet (2012: 1444). Es común pensar que con la Internet sobrevino lo que se conoce como explosión de la información. Sin embargo, Wilson, en su recuento de los cincuenta años de investigación en conducta de información, habla de una explosión de la información que empieza a gestarse durante la Segunda Guerra Mundial: ambos bandos buscaban mejorar su maquinaria bélica, trayendo como resultado una gran cantidad investigación científica y tecnológica que dio lugar a gran cantidad de reportes de acceso restringido. Al final de la guerra, esa información, principalmente proveniente de Alemania, estuvo disponible a la comunidad científica de los aliados. Es para los años sesenta que se puede hablar de una explosión de la información, habida cuenta de la amplia disponibilidad de información, antes secreta, y los múltiples programas de uso pacífico de la energía atómica generaron numerosas investigaciones que fueron publicadas (2009: 27-28).

La recuperación de información es un tema bastante amplio y complejo, donde confluyen varias disciplinas, cada una con sus aportes de investigación en el tema a lo largo de todos estos años, sin embargo, esa variedad de disciplinas ha hecho imposible tener un concepto único de recuperación de información, común a todas las que la estudian (Salvador 2008: 9). Fue Calvin N. Mooers quien acuñó el término “*information retrieval*” en 1951, según se puede apreciar en la cita hecha por Saracevic de la definición que hizo Mooers de recuperación de información. A continuación la cita traducida:

Recuperación de información es el nombre para el proceso o método donde un posible usuario de información es capaz de convertir su necesidad de información en una lista real de citas de documentos almacenados y que contienen información útil para él (...) Recuperación de información abarca los aspectos intelectuales de la descripción de la información y su especificación para la búsqueda, y también cualquier sistema, técnica, o máquina que se utilice para llevar a cabo la operación.". (Mooers citado por Saracevic 2012: 51)

A pesar de haber transcurrido más de sesenta años de esta definición, bien dice Saracevic, la idea formulada por Mooers es aún válida. Aunque no se menciona la palabra *relevante*, ésta está implícita en la palabra *útil*. (2012: 51).

En la perspectiva histórica de las investigaciones en recuperación de información de Sanderson y Croft es interesante ver que los sistemas de recuperación de información han ido creciendo en complejidad, en la medida que la cantidad de información disponible a los usuarios aumentaba, con el surgimiento de varias propuestas para mejorar la relevancia mediante algoritmos y métodos (2012: 1444-1450). Para el tema de recuperación de información es muy importante el concepto de relevancia. Mizzaro hace un estudio bastante exhaustivo de la literatura existente en la época sobre la relevancia, indicando que no hay y no ha habido consenso para definir la relevancia, principalmente porque ésta ha sido estudiada desde varias disciplinas (1997: 810-828). En su artículo sólo limita su estudio a la ciencia de la información, la documentación y la recuperación de la información. Se trata básicamente de un análisis histórico de artículos publicados hasta entonces sobre el tema. Para Borlund, la relevancia tiene carácter multidimensional y dinámica, y es justo eso lo que explicaría que no haya aún un consenso para llegar a definirla (2003: 913). Es multidimensional porque es percibida y juzgada de manera distinta por cada usuario. Es dinámica porque la percepción del usuario puede cambiar en el tiempo (2003: 914). La más reciente publicación de Saracevic es una perspectiva histórica de la investigación sobre la

relevancia dentro de la ciencia de la información, donde dice que la fortaleza de la relevancia es su carácter intuitivo porque todos pueden comprenderla intuitivamente sin mayores explicaciones, pero su debilidad es ser una noción humana, a la cual todos los algoritmos tratan de acercarse, y no técnica, lo cual la hace difícil de controlar y comprender (2012: 49). En la bibliografía usada por Sanderson y Croft se puede ver claramente que a lo largo de estos últimos cincuenta años se han ido desarrollando y proponiendo mejoras de relevancia para los tres principales modelos de recuperación de información: booleano, vectorial y probabilístico (2012: 1446-1449).

1.3. Los catálogos en línea

Los catálogos en línea, comúnmente conocidos como *OPAC (Online Public Access Catalogue)* en el ámbito bibliotecológico, aparecieron a finales de los años setenta y principios de los ochenta (Beheshti 1997: 163). Estos constituyen sistemas de recuperación de información, cuya primera generación fue una extensión de los tradicionales ficheros bibliográficos que ofrecían tres puntos de acceso: título, autor y tema. Esta vendría a ser la primera generación de *OPAC*, con limitaciones para mostrar información pero a pesar de ello empezó a ganar popularidad sobre el catálogo de fichas (1997: 164). La segunda generación de *OPAC* fue marcada por el uso de búsquedas booleanas, operadores de proximidad, exploración de registros según el ordenamiento de estos en los estantes, la búsqueda por palabra clave y muchas mejoras en la manera de mostrar la información recuperada (1997: 164). Aún hoy en día sigue predominando el modelo booleano, a pesar de su antigüedad y de las alternativas que han ido apareciendo. Al respecto, Cooper manifiesta que no ha habido un consenso claro entre los investigadores para determinar qué modelo alternativo es mejor que el booleano. Detalla cinco problemas del modelo booleano y llega a proponer la solución respectiva

para cada uno de ellos (1988: 265-267). Uno de esos problemas indica que este modelo genera resultados vacíos o excesivos (1988: 265). Para el catálogo en línea objeto de este estudio se usa el método booleano para la recuperación de información, y la frecuencia de términos para hacer el cálculo de la relevancia. Más adelante se dará más detalles del análisis del motor de búsqueda del catálogo en cuestión. La tercera generación de *OPAC* se vio beneficiada por el desarrollo del modelo cliente/servidor y los protocolos TCP/IP y Z39.50, permitiendo el acceso remoto a otros catálogos desde la misma interfaz de búsqueda (Beheshti 1997:165-166).

La llegada de la Web marcó profundamente el desarrollo y evolución de los *OPAC* en la segunda mitad de los noventa. Para Babu y O'Brien en esta etapa aparecieron los *OPAC* avanzados con mejoras sustanciales en la interfaz gráfica y usabilidad, facilitando la navegación entre registros mediante el hipertexto, permitiendo enlazar las referencias al texto completo (2000: 317).

1.4. El análisis de las bitácoras de transacciones

1.4.1. Definición

El tema del análisis de bitácoras de transacciones data, según la literatura revisada, desde la segunda mitad de los años sesenta, cuando aparece una publicación sobre un proyecto llevado a cabo por David Meisser y D. J. Sullivan, el cual consistió en evaluar las reacciones de los usuarios a un prototipo de sistema de recuperación de información en línea (Peters 1993: 42). Desde entonces, y hasta la fecha, se han publicado los proyectos de investigación llevados a cabo en distintas áreas, especialmente en la de bibliotecología y ciencia de la información.

Es pertinente definir lo que es una bitácora de transacciones en esta etapa inicial del presente trabajo. Una transacción se define como una acción iniciada por un usuario (del sistema). Para el tema que aquí se trata, la acción iniciada por el usuario y la respuesta del sistema constituyen juntas una única transacción. También es válido que el sistema inicie una acción y que el usuario dé una respuesta. Los sistemas de recuperación de información tienen mecanismos que permiten capturar toda la información que inicia (consulta) y concluye (respuesta) una transacción. Estos mecanismos almacenan toda esa información en un archivo de texto denominado *bitácora de transacciones*. El análisis de las bitácoras de transacciones se puede definir como el estudio de las interacciones registradas electrónicamente entre los sistemas de recuperación de información en línea y las personas que buscan la información contenida en dichos sistemas (Peters y otros 1993: 39). En el caso concreto de la presente investigación se refiere a bitácoras de transacciones de búsquedas de un catálogo en línea.

1.4.2. Dificultades de su uso

La cantidad y tipo de información que almacenan las bitácoras de transacciones dependen de la naturaleza del sistema que los genera. La estructura de estas bitácoras depende del fabricante del software. Sandore y otros proponen un manifiesto donde recomiendan la estandarización de las bitácoras de transacciones de los sistemas de recuperación de información, habida cuenta que su análisis requería un trabajo bastante intensivo de preparación de los datos. Además, se propone la creación de software que facilite el análisis mediante herramientas amigables al alcance de los bibliotecarios (1993: 105-106). Casi veinte años después se puede asegurar que el tan esperado estándar no ha llegado a concretarse. Las herramientas que permiten el análisis de éstas

sí han mejorado pero aun así requieren un conocimiento de nivel intermedio/avanzado para procesar toda esa información de manera adecuada. La preparación de los datos, previa al análisis, sí requiere de software o rutinas de programación un tanto sofisticadas. No todos los investigadores interesados en estos temas tienen a su alcance personas capacitadas que les faciliten los datos en un formato de fácil análisis. Esto puede estar desanimando la investigación en esta área.

Contar con un estándar sería de gran ayuda para el estudio y análisis de estas bitácoras porque se podrían usar metodologías y modelos de estudios ya probados y exitosos en cualquier sistema sin importar el origen y/o fabricación del software. Para esta investigación, la bitácora de transacciones debe contener los términos de búsqueda empleados por el usuario para consultar al sistema, además de información como tipo de búsqueda, filtros usados, número de registros recuperados, entre otros. No todos los sistemas de recuperación de información registran los mismos datos en las bitácoras de búsquedas. Por esa razón, habrá casos donde no se podrá aplicar todos los métodos propuestos en la literatura revisada, como los que propone Kaske (1993: 79-86).

1.4.3. Tipos de análisis

En el ámbito de la bibliotecología y ciencia de la información, la mejor revisión bibliográfica encontrada con respecto a la historia del estudio del análisis de transacciones es el artículo de Peters (1993). El autor divide el estudio del análisis de transacciones en tres fases: la primera, de mediados de los años sesenta hasta fines de los años setenta, donde se puso énfasis en evaluar el desempeño de los sistemas más que el comportamiento y desempeño del usuario. La segunda fase va desde fines de los años setenta hasta mediados de los años ochenta. Aquí fue la primera vez que el análisis de bitácoras de transacciones se aplicó al estudio de los catálogos en línea. En

general, los investigadores estuvieron interesados de igual manera en saber cómo se usaban los sistemas (por ejemplo, qué opciones de búsqueda se escogían) y en la investigación del comportamiento de los usuarios (por ejemplo, cuánto duraba en promedio una sesión de búsqueda, cuántos y qué tipos de errores se cometían). La tercera fase va desde mediados de los años ochenta hasta 1993 (año en que se escribió el artículo) y se ha caracterizado por la diversificación. Algunos se concentraron en investigar aspectos específicos de búsquedas (por ejemplo, búsqueda por temas), grupos específicos de usuarios, u otros tipos de sistemas de información (por ejemplo, estaciones de trabajo de CD-ROM), u otros tipos de bases de datos. Réplicas de estudios previos y estudios longitudinales (investigaciones que implican observaciones repetidas de los mismos objetos de estudio durante largos períodos de tiempo) empezaron a aparecer. La revisión que hace el autor está organizada por el tipo de estudio realizado en orden cronológico (Peters 1993: 41-66). No se ha encontrado un estudio similar al de Peters para los años posteriores a 1993 y hasta la fecha, donde es un hecho que la Internet, la Web 2.0 y la popularidad de los dispositivos móviles hacen necesario y muy útil el análisis de bitácoras de transacciones (Church y otros 2007: 1-3).

Dentro de los tipos de análisis se han estudiado las situaciones donde la recuperación de información es vacía, como de la que es objeto el presente trabajo de investigación. Aquellos sistemas que usan el método de coincidencia más cercano (a la palabra o frases usadas en las búsquedas) quedan excluidos de este tipo de estudio porque no dan resultados vacíos. Los sistemas que usan este método siempre encuentran resultados, a pesar de que los usuarios formulan sus búsquedas con palabras mal escritas o con palabras o frases incompletas. Detrás de esto está lo que se denomina la lógica difusa (*fuzzy logic*), lo cual permite corregir las palabras mal escritas o parciales para obtener resultados que responden a las consultas de los usuarios.

1.4.4. Estudios de uso de catálogos en línea mediante el análisis de bitácoras de transacciones

El análisis de bitácoras de transacciones es una metodología que se viene usando desde hace varias décadas para poder medir y determinar el comportamiento humano frente a los sistemas de recuperación de información, los cuales tienen un mecanismo que permite registrar imperceptiblemente, en archivos de texto, la interacción entre el usuario y el sistema. Aquí las posibilidades de análisis son muy variadas y dependen de lo que estas bitácoras registran. Centrándose en el tema de estudio del presente trabajo, se revisarán principalmente estudios que analizan la recuperación nula o vacía en catálogos en línea.

Dickson considera que las búsquedas sin resultados se deben a errores que los usuarios cometen influenciados por modelos conceptuales generados de experiencias previas con otros catálogos (1984: 35). El usuario puede formular muy bien sus términos de búsqueda y obtener un resultado vacío. Los errores también pueden deberse a la calidad de los datos del catálogo: mal ingreso de datos, omisión de datos, errores de escritura, entre otros. Lo que se desea saber es cuánto peso tienen los errores cometidos por los usuarios en relación a los cometidos por la biblioteca. Dickson hace un análisis de una muestra de búsquedas con resultados vacíos pero solamente para búsquedas por título y por autor (1984: 26-35). Su objetivo era encontrar la razón de las fallas de las búsquedas y entender el modelo conceptual de los usuarios del catálogo en línea de la biblioteca de la Universidad de Northwestern. Parte de la hipótesis de que hay una diferencia entre el modelo del sistema que tiene el usuario con respecto a las capacidades del catálogo en línea que empieza a usar. Manifiesta que ésta es probablemente la causa de muchos de los errores de los usuarios.

En un trabajo posterior al de Dickson, Peters reporta un estudio que busca determinar la proporción de fallas, el uso de patrones, y los problemas de uso del catálogo en línea de la Universidad de Missouri-Kansas City (1989: 267-273). Como proporción de fallas se refiere a aquellas búsquedas que no obtienen resultados. Sin embargo, aclara que no está de acuerdo con la definición dada por Cheryl Kern-Simirenko sobre lo que son búsquedas exitosas o búsquedas fallidas en los catálogos en línea. Para Kern-Simirenko una búsqueda fallida era aquella que u obtenía cero resultados u obtenía resultados sin sentido. Una búsqueda exitosa era aquella que obtenía uno o más resultados (1983: 27). Efectivamente, tal como lo manifiesta Peters, no se puede estar de acuerdo con las definiciones de Kern-Simirenko. Un usuario puede formular bien su búsqueda y recibir como respuesta, por parte del sistema, que la biblioteca no tiene el ítem que busca entre sus fondos. Peters propone tipificar las búsquedas fallidas: errores del usuario (cuando el usuario falla en encontrar un ítem que sí se encuentra en la biblioteca), fallas de la colección (cuando la colección no contiene el ítem buscado), y fallas de alcance (cuando los proyectos de conversión retrospectiva de registros bibliográficos a nuevos formatos no se han completado, o no se da soporte a ciertos formatos) (1989: 267-273). En el análisis de las probables causas de los problemas de los usuarios en las búsquedas, llevado a cabo en el catálogo de la Universidad de Missouri-Kansas City, se llegan a determinar catorce tipos, entre los que destacan los errores tipográficos y ortográficos. En general, considerándose todos los tipos de búsqueda, la proporción de errores que llega a registrarse fue de 47.6%.

Zink presenta el caso del catálogo en línea de la Universidad de Nevada donde solo se analiza la bitácora de transacciones de veinticuatro horas de uso del catálogo debido a la gran cantidad de datos que significa analizar periodos más extensos (1991: 49-56). Su objetivo era analizar las búsquedas no exitosas para cada índice, palabra por

palabra: título, autor, temas y otros, siendo la búsqueda por título el tipo de búsqueda menos exitosa con 41%. En total, el 27% de las búsquedas fueron no exitosas. Aunque no llega a dar una definición de búsqueda no exitosa, se desprende de la descripción de su análisis que es aquella que falla en recuperar al menos un registro bibliográfico. No llega a tipificar los errores como lo hizo Peters (1989: 270) sino a dar razones, a las cuales les asignó porcentajes, aunque se puede establecer cierta similitud entre los tipos de errores de Peters y las razones de búsquedas no exitosas de Zink (1991), quien además manifiesta que una de las causas de los errores cometidos por los usuarios en las búsquedas por autor tenía que ver con la costumbre adquirida del uso del anterior catálogo en línea, lo cual tiene cierta similitud en la afirmación antes mencionada de Dickson con respecto al modelo conceptual adquirido previamente por el usuario en otros sistemas como causa de los muchos errores cometidos por los usuarios.

Hacia fines de los años noventa, Hildreth reporta un estudio sobre las búsquedas booleanas y por palabra clave en el catálogo en línea de la Universidad de Oklahoma, para el cual se hizo uso de cuestionarios y análisis de bitácoras de transacciones con el fin de complementarlos (1997: 52-63). Se deseaba descubrir si los usuarios tenían cierto nivel de comprensión con respecto a cómo el sistema procesaba las búsquedas booleanas y por palabra clave. De igual manera se deseaba saber qué tan frecuente era el uso de las búsquedas por palabra clave en comparación con las búsquedas por título, autor y tema. Se llegó a determinar que el tipo de búsqueda más usado era por palabra clave con 30.6% frente a 30.5% del tipo título. Presenta una tabla de rangos por número de registros recuperados: 0, 1-89 y 90 a más. Siendo la búsqueda por título la de mayor porcentaje con cero resultados: 38.5%. Por palabra clave, el 27% de búsquedas no mostraba resultados.

Lau y Goh llevan a cabo un estudio en la Universidad Tecnológica de Nanyang (Singapur) donde analizan la bitácora de búsquedas correspondiente a un semestre académico (2006: 1316-1329). Se plantearon dos objetivos: 1) descubrir los patrones en las conductas de búsqueda de los usuarios, y 2) saber las razones de las búsquedas fallidas. Se investigó lo siguiente para poder cumplir con los objetivos: longitud promedio, mínima y máxima, de las consultas, frecuencia y tipos de opciones de búsqueda usados, frecuencia y tipos de operadores booleanos usados, tipos de errores tipográficos, relación entre las búsquedas fallidas y las opciones de búsqueda, relación entre las búsquedas fallidas y los operadores booleanos, y la relación entre las búsquedas fallidas y la longitud de las búsquedas. El tipo de búsqueda más usado fue por palabra clave con 68.9% y llegó a registrar el 32.21% de respuestas sin resultados, el más alto de todos. En total, el porcentaje de búsquedas sin resultados fue de 49.50%. El operador booleano más usado fue AND (10.28% del total de búsquedas realizadas) y también fue el que más resultados vacíos obtuvo (6.01% del total de resultados vacíos). Los operadores booleanos combinados más usados fueron OR y AND (1.37%). Esta combinación fue la que más resultados vacíos obtuvo, llegando a registrar el 0.64% del total de resultados vacíos. Ante este panorama, los autores llegan a proponer la reformulación interactiva de las consultas como una técnica de asistencia a los usuarios que recomienda alternativas de consultas. También proponen la navegación o exploración (*browsing*) para mejorar el pobre desempeño de los encabezamientos de materia. Finalmente propone la ayuda contextual que incluye recomendaciones de términos y documentos, ayuda personalizada, y enlaces a expertos y usuarios con intereses similares. Todo esto con la ayuda de perfiles de usuarios con la consiguiente necesidad de registrarse en el sistema antes de realizar las búsquedas.

Ortiz-Repiso y otros realizaron una investigación sobre el uso del catálogo en línea de la red de bibliotecas del Consejo Superior de Investigaciones Científicas de España mediante el uso de cuestionarios y complementados con el análisis de bitácoras de transacciones (2006). Los cuestionarios se usaron para determinar aspectos demográficos como áreas de especialización, preferencias de uso bases de datos bibliográficas, y con respecto al catálogo en línea, lugares desde donde lo usan, maneras de aprender su uso, grados de satisfacción con él y facilidad de uso, utilidad de las ayudas, comprensión de la terminología, y preferencias por tipo de búsqueda, entre otros (2006: 190-211). Los resultados se presentan en tablas y gráficos, sin embargo en ninguna de ellas se relacionan o comparan con aquellas provenientes del análisis de las bitácoras, a diferencia de Milliari y Kyriaki-Manessi que aplicaron un método más interesante para usar un cuestionario y bitácoras de transacciones complementarios en la biblioteca de la Universidad de Macedonia (2007: 107-122). Implementaron un formulario en línea que aparecía antes de iniciar la búsqueda en el catálogo. De esa manera podían relacionar todas las acciones hechas por los usuarios con sus categorías (pregrado, posgrado, etc.), departamento, año de ingreso, rango académico; con los cuales podían determinar si el nivel académico tenía relación con la efectividad de las búsquedas, lo cual fue confirmado por el estudio. Aunque no definieron lo que es una búsqueda fallida ni tampoco se dan cifras que indiquen a qué grado llegan las fallas, reportaron que la búsqueda por tema es aquella donde los usuarios cometen más fallas, seguida de la búsqueda por ISBN, serie, y palabra clave básica. En cambio, Villen-Rueda, Senso y de Moya realizaron un análisis de bitácoras de transacciones de las búsquedas por tema de la Universidad de Granada para establecer la relación entre los tipos de usuarios y los tipos de búsqueda, sin necesidad de cuestionarios, para identificar los tipos de usuarios (2007: 327-337). En su lugar agruparon los números IP

de las bitácoras según su procedencia; así, podían llegar a establecer que determinadas transacciones de búsquedas provenían de profesores, alumnos o bibliotecarios, por ejemplo. Los investigadores encontraron que la frecuencia de búsquedas por navegación es similar a los otros tipos de búsqueda. Solo el 14% de las búsquedas se direcciona al índice de tema contra el 49% de búsquedas por títulos, y 37% de búsquedas por autor. No se hizo análisis de búsquedas fallidas o con resultados nulos. En la misma línea de interés por las búsquedas temáticas, Antell y Huang llevaron a cabo un estudio que complementaba entrevistas con análisis de bitácoras para determinar el grado de éxito en las búsquedas por tema, profundizando en las razones de las fallas en este tipo de búsqueda (2008: 68-76). Esta metodología les permitió observar que no existía correlación entre los temas hallados por los usuarios y su propia percepción de búsquedas temáticas exitosas. De todas las búsquedas analizadas, solo el 4.6% correspondía a búsquedas por tema, con un porcentaje de 48.8% de resultados vacíos.

Moulaison realizó un análisis para determinar patrones de consultas en el catálogo de la biblioteca de *The College of New Jersey* (2008: 230-237). En las búsquedas mayormente se usaron de una a tres palabras. El porcentaje de búsquedas sin resultados llegó a 30%. Resulta interesante el estudio de las estrategias aplicadas por los usuarios ante las búsquedas sin resultados: el 52.3% modificó sus términos y el 35.9% cambió de índice.

Regresando a la hipótesis que planteó Dickson en 1984 con respecto a la diferencia que existe entre el modelo conceptual que tiene un usuario por su experiencia previa y el nuevo modelo que representa el catálogo en línea como la causa de muchos de los errores que cometen en las búsquedas, es pertinente mencionar estudios recientes llevados a cabo en tres universidades de la India para determinar la influencia en los usuarios de las búsquedas que realizan en la Web al momento de usar los catálogos en

línea de las bibliotecas. Aunque aquí no se llevó a cabo un análisis transaccional de las bitácoras de búsquedas sino la aplicación de cuestionarios a quinientos usuarios entre académicos, investigadores y estudiantes de posgrado para determinar sus costumbres con respecto a sus destrezas computacionales y nivel de uso de búsquedas en la Web, así como también el uso del catálogo en línea y sus preferencias, actitudes y expectativas. A pesar del uso extendido de la Web, poco más del 40% de los entrevistados declaró hacer uso frecuente o muy frecuente del catálogo (Kumar 2011: 14-21).



CAPÍTULO 2: EL ENTORNO

2.1. El Sistema de Bibliotecas de la PUCP

Desde la fundación de la Universidad en 1917, la biblioteca empezó a brindar sus servicios a los primeros estudiantes y profesores. Desde entonces fue adecuándose a las exigencias que planteaba el crecimiento de la Universidad hasta llegar a ser lo que es hoy en día: un sistema de bibliotecas compuesto por una Biblioteca Central y siete bibliotecas departamentales. Adicionalmente, presta servicios a los usuarios de las bibliotecas de la Escuela de Negocios CENTRUM, del Instituto Confucio y del Instituto Riva-Agüero. Estas no dependen administrativamente de la Biblioteca Central pero sí comparten el mismo software de gestión de bibliotecas, lo cual les exige la adhesión a las políticas de catalogación que sigue el actual Sistema de Bibliotecas.

2.1.1. Servicios brindados

A través de las nueve bibliotecas que conforman el sistema se brinda el servicio de préstamo de materiales bibliográficos que está condicionado por el tipo de material y por el tipo de usuario. Se da prioridad a los alumnos matriculados y a los profesores, quienes tienen préstamos a domicilio. Por razones de integridad y conservación de las colecciones, solo se permite la consulta en sala de las revistas, las colecciones especiales, los recursos audiovisuales, los materiales de referencia, y las tesis de los exalumnos de la universidad.

Desde el primer semestre del 2010 se empezó a ofrecer el servicio de estantería abierta en la Biblioteca Central, en el segundo y tercer pisos, aunque a manera de prueba piloto, con la finalidad de afinar los detalles y empezar a ofrecer plena y oficialmente el servicio en el segundo semestre del 2010. De esta manera se puso en servicio de estantería abierta alrededor de 400 mil volúmenes. Este hecho tuvo un

impacto importante en el uso del catálogo en línea. Según las cifras recogidas al final del 2010 (ver Anexo 12), el número de búsquedas bajó en un 23.78% con respecto al año anterior. El 2011 el porcentaje de búsquedas se incrementó ligeramente en 1.78% (ver Anexo 13) y el 2012, 2.04% (ver Anexo 14).

El préstamo inter bibliotecario está disponible para los usuarios cuando estos encuentran materiales que no existen en sus colecciones pero sí en las instituciones con las que mantiene convenios para tal fin.

Para facilitar la devolución de materiales se ha dispuesto el funcionamiento de buzones de auto devolución en la zona externa de las bibliotecas que tienen mayor flujo de sus colecciones.

En cuanto a colecciones digitales, se cuenta con varios paquetes de bases de datos de prestigio y calidad, los cuales constituyen una colección electrónica de artículos de revistas de grandes dimensiones y que cubren todas las disciplinas que se imparten en la Universidad. Para la búsqueda centralizada de todos los recursos electrónicos contratados y aquellos de acceso abierto, previamente seleccionados, se ha puesto en funcionamiento un sistema de búsqueda y descubrimiento, desde donde también es posible encontrar referencias a los registros bibliográficos del catálogo en línea.

2.1.2. La colección bibliográfica

La colección bibliográfica reúne a más de 700 mil volúmenes, entre monográficos, publicaciones periódicas, recursos electrónicos, audiovisuales, multimedia, software, tesis, entre otros. La cobertura temática abarca todas las disciplinas impartidas en la Universidad, siendo las más numerosas las de derecho, ciencias e ingeniería, ciencias sociales y humanidades.

Todas las colecciones son catalogadas según las Reglas de Catalogación Angloamericanas 2 y se usa el formato MARC21 para la creación de registros bibliográficos. El esquema de clasificación predominante es el de la Biblioteca del Congreso de los Estados Unidos de Norteamérica (*Library of Congress*). El esquema de clasificación Dewey se usa en las colecciones de derecho y las del Instituto Riva-Agüero. También se usan esquemas de clasificación locales para algunas colecciones como las de tesis y las del Centro de Documentación de Ciencias Sociales.

Para la indexación temática se usan los *encabezamientos de materia* de la Biblioteca del Congreso (*Library of Congress*). Con respecto a las autoridades, solo se ha automatizado el control de los nombres de personas como autores o como tema, a través del sistema *Symphony*. Existe un listado de autoridades temáticas que aún no se ha automatizado y que los bibliotecarios usan al momento de indizar sus registros bibliográficos.

2.1.3. Los usuarios

Los usuarios del Sistema de Bibliotecas de la PUCP son todos los miembros de la comunidad universitaria, principalmente alumnos y docentes. Según cifras del semestre 2012-2, el número de alumnos de pre y posgrado llegó a 24,482. El número de docentes llegó a 2,380 frente 2,710 correspondiente a personal administrativo. Los egresados también reciben servicios.

Según cifras del 2012, la circulación anual fue de 1'894,629 consultas de material bibliográfico.

2.2 El catálogo en línea

En diciembre de 1997 se inició el servicio de catálogo en línea a través de la interfaz llamada WebCat® que, desde su aparición, estuvo en constante renovación tanto en diseño como en funciones. En febrero del 2006 fue reemplazado por iLink®, interfaz de última generación para la época que permitía integrar contenidos enriquecidos como cubiertas de libros, reseñas, resúmenes, tablas de contenido, etc. Los usuarios pueden reservar el préstamo de materiales que se encuentran prestados por otros. De igual manera, es posible renovar los préstamos no vencidos. A cada usuario se le asigna una cuenta desde la cual puede gestionar los servicios que se acaban de mencionar, además de poder consultar el historial de sus préstamos, crear listas bibliográficas, y gestionar sus autores y temas favoritos.

En marzo del 2009, se implementó la interfaz e-Library®, que incluye adecuaciones para dispositivos móviles. Las interfaces antes mencionadas son de la empresa *SirsiDynix Corp.*, proveedora del software integrado para la automatización del Sistema de Bibliotecas PUCP.³

2.2.1 Características y funciones

e-Library® es una interfaz gráfica de usuario que ofrece acceso al catálogo bibliográfico y a otros servicios complementarios. Permite integrar contenidos a los registros bibliográficos y compartir información en las redes sociales. Está lista para ser

³ El Sistema de Bibliotecas de la PUCP usa el sistema integrado para automatización de bibliotecas Symphony, de la empresa SirsiDynix, desde el año 1997. Este software se llamó Unicorn desde que salió al mercado en 1983. En junio del 2005, la empresa Sirsi Corp. anuncia su fusión con la empresa Dynix corp. llegando a formar SirsiDynix Corp. Producto de esa fusión, en el año 2007, sale al mercado Symphony, que en la práctica viene a ser Unicorn con las mejores funcionalidades de Horizonte, producto similar que ofrecía Dynix Corp. antes de la fusión.

usada por dispositivos móviles desde la Web y también puede ser consultada desde la aplicación móvil BookMyne®, con funciones similares a las de e-Library®.

Las características y funciones que se describirán a continuación corresponden solo a aquellas que se han configurado para el Sistema de Bibliotecas de la PUCP:

- Búsquedas booleanas a través de dos tipos de búsqueda: 1) simple y 2) avanzada. Ambas tienen opciones complementarias para refinar los resultados.
- Búsqueda por código de clasificación que permite localizar los ítems en los estantes (signatura topográfica).
- Búsquedas en catálogos especiales para facilitar la localización de tesis, artículos de revistas y material audiovisual. En todos estos casos la búsqueda es del tipo simple.
- Búsquedas en catálogos externos previamente configurados mediante el protocolo Z39.50.
- Categorización de resultados de búsqueda según código de clasificación asignado a los registros. Esto permite filtrar los resultados y contextualizarlos a áreas temáticas asociadas a los códigos de clasificación de *Library of Congress* o *Dewey*.
- Durante la sesión de búsqueda se guardan los términos empleados en búsquedas anteriores a la actual con la finalidad de reutilizarlos.
- Cualquier página del catálogo, incluidos los registros bibliográficos y resultados de búsquedas pueden guardarse en la lista de favoritos del navegador. También es posible crear enlaces permanentes a estas páginas.
- Configuración de enlaces a páginas sobre los servicios de las bibliotecas.
- Reserva de material bibliográfico desde los resultados de búsquedas.
- Renovación de préstamos no vencidos.

- Información sobre préstamos actuales, reservas, multas, suspensiones e historial de préstamos a través de Mi Cuenta. También es posible cancelar reservas desde aquí.
- Lista de autores y temas favoritos según preferencias del usuario con posibilidad de alertas por correo electrónico cuando se han ingresado nuevos materiales al catálogo de estos favoritos.
- Creación de listados bibliográficos personalizados.

2.2.2 El motor de búsqueda

BRS/Search es una base de datos de texto completo y un sistema de recuperación de información. BRS/Search utiliza un sistema de indexación completamente invertido para almacenar, localizar y recuperar datos no estructurados (Wikipedia 2013). Este subyace en e-Library® y constituye su motor de búsqueda.

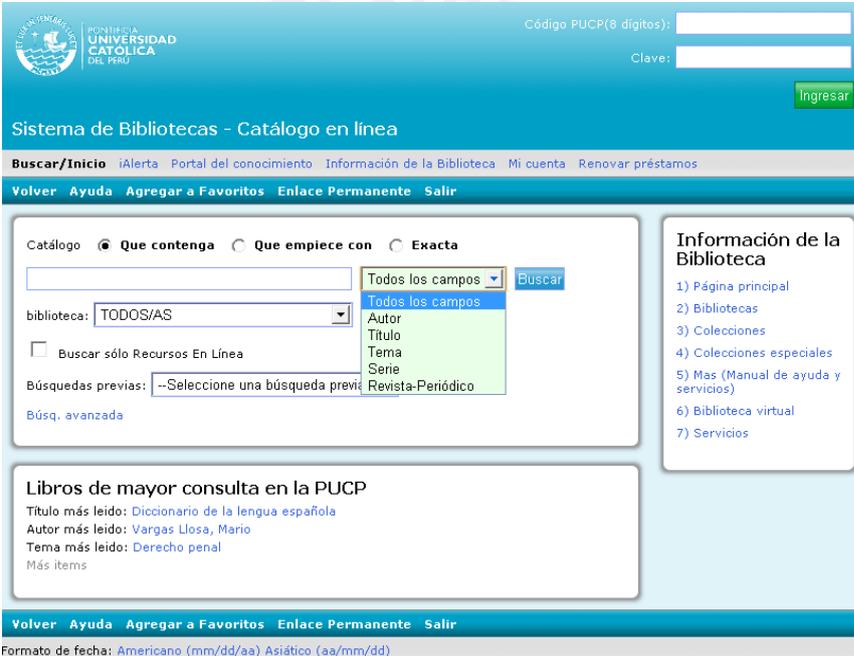
Se puede hacer búsquedas de palabras únicas o de frases, en campos específicos o en más de un campo mediante los operadores booleanos AND, OR, NOT y XOR para combinar términos de búsqueda. También permite las búsquedas de términos adyacentes usando los operadores de posición SAME, WITH, NEAR y ADJ. Se incluyen otras opciones como el truncamiento tanto hacia la derecha como hacia la izquierda usando el signo "\$", y operadores relacionales como >, <, =, >=, <=, <> para valores numéricos. El signo de interrogación "?" se usa para la sustitución de caracteres.

Entre los tipos de búsquedas, ofrece:

- Búsqueda simple: es la búsqueda por defecto y dirige todos los requerimientos al índice general si el usuario no selecciona otro índice antes de enviar su requerimiento de búsqueda. En este están indizados todos los campos de los registros bibliográficos. Ofrece tres opciones:

- Que contenga: permite localizar y mostrar registros que contienen la palabra o frase usada en la búsqueda.
- Que empiece con: genera un listado alfabético con términos similares a la raíz del término usado en la búsqueda. (Imagen 2)
- Exacta: es una búsqueda por palabra clave en los índices de los encabezamientos.

Dentro de la búsqueda simple es posible usar operadores booleanos.



The screenshot shows the 'Sistema de Bibliotecas - Catálogo en línea' interface. At the top right, there are input fields for 'Código PUCP(8 dígitos):' and 'Clave:', with an 'Ingresar' button. Below this is a navigation bar with links: 'Buscar/Inicio', 'Alerta', 'Portal del conocimiento', 'Información de la Biblioteca', 'Mi cuenta', and 'Renovar préstamos'. A secondary navigation bar contains: 'Volver', 'Ayuda', 'Agregar a Favoritos', 'Enlace Permanente', and 'Salir'.

The main search area features three radio buttons for search type: 'Catálogo' (selected), 'Que empiece con', and 'Exacta'. Below these are a search input field, a dropdown menu for search scope (currently showing 'Todos los campos'), and a 'Buscar' button. A 'biblioteca:' dropdown is set to 'TODOS/AS'. There is a checkbox for 'Buscar sólo Recursos En Línea' and a 'Búsquedas previas:' section with a selection prompt. A link for 'Búsq. avanzada' is also present.

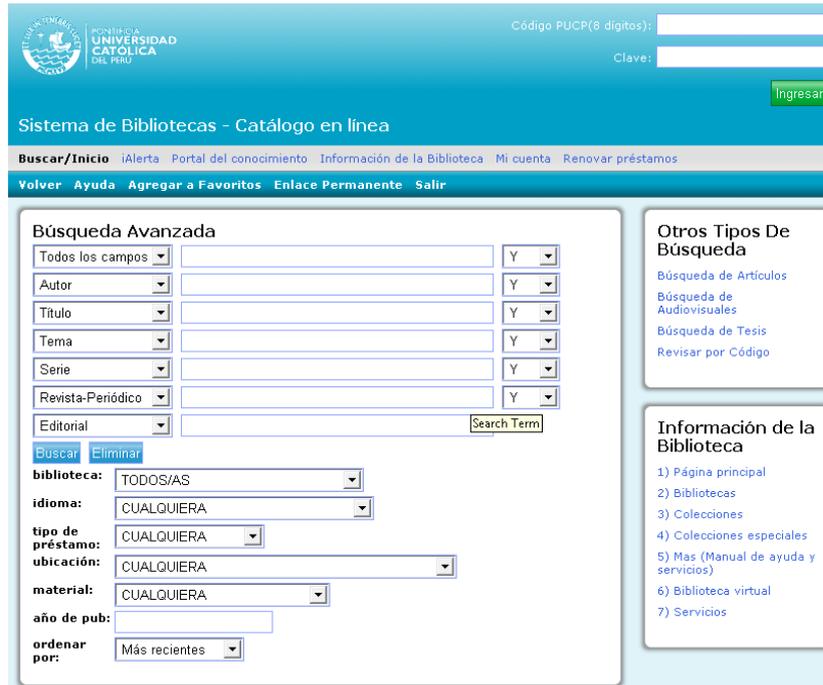
On the right side, there is a box titled 'Información de la Biblioteca' with a numbered list of links: 1) Página principal, 2) Bibliotecas, 3) Colecciones, 4) Colecciones especiales, 5) Mas (Manual de ayuda y servicios), 6) Biblioteca virtual, and 7) Servicios.

Below the search form is a section titled 'Libros de mayor consulta en la PUCP' listing: 'Titulo más leído: Diccionario de la lengua española', 'Autor más leído: Vargas Llosa, Mario', and 'Tema más leído: Derecho penal', with a 'Más items' link.

At the bottom, there is another navigation bar with 'Volver', 'Ayuda', 'Agregar a Favoritos', 'Enlace Permanente', and 'Salir', and a footer note: 'Formato de fecha: Americano (mm/dd/aa) Asiático (aa/mm/dd)'.

Imagen 1: Pantalla del formulario de búsqueda simple

- Búsqueda avanzada: permite combinar los términos de búsqueda en índices específicos a través de operadores booleanos con la finalidad de obtener resultados más precisos. Adicionalmente, se puede aplicar filtros como tipo de material, ubicación, biblioteca, tipo de préstamo y año de publicación.



Sistema de Bibliotecas - Catálogo en línea

Código PUCP(8 dígitos):
Clave:

[Buscar/Inicio](#) | [Alerta](#) | [Portal del conocimiento](#) | [Información de la Biblioteca](#) | [Mi cuenta](#) | [Renovar préstamos](#)

[Volver](#) | [Ayuda](#) | [Agregar a Favoritos](#) | [Enlace Permanente](#) | [Salir](#)

Búsqueda Avanzada

Todos los campos Y
 Autor Y
 Título Y
 Tema Y
 Serie Y
 Revista-Periódico Y
 Editorial

biblioteca:
 idioma:
 tipo de préstamo:
 ubicación:
 material:
 año de pub:
 ordenar por:

Otros Tipos De Búsqueda

[Búsqueda de Artículos](#)
[Búsqueda de Audiovisuales](#)
[Búsqueda de Tesis](#)
[Revisar por Código](#)

Información de la Biblioteca

- 1) Página principal
- 2) Bibliotecas
- 3) Colecciones
- 4) Colecciones especiales
- 5) Mas (Manual de ayuda y servicios)
- 6) Biblioteca virtual
- 7) Servicios

Imagen 2: Pantalla del formulario de búsqueda avanzada



Sistema de Bibliotecas - Catálogo en línea

[Buscar/Origen](#) | [Alerta](#) | [Portal del conocimiento](#) | [Información de la Biblioteca](#) | [Mi cuenta](#) | [Renovar préstamos](#)

[Regresar](#) | [Ayuda](#) | [Nueva búsqueda](#) | [Anterior](#) | [Siguiente](#) | [Salir](#)

ALL Búsqueda alfabética de catálogo por materia: "recuperacion de informacion"

1.	Recuerdo de dos lunes--Crítica e interpretación Miller, Arthur, 1915-2005.	2
2.	Recuperación de la información	44
3.	Recuperación de la información--Aplicaciones de sistemas expertos	1
4.	Recuperación de la información--Congresos--Gran Bretaña	1
5.	Recuperación de la información--Encuestas	1
6.	Recuperación de la información--Estudio y enseñanza	4
7.	Recuperación de la información--Evaluación	3
8.	Recuperación de la información multilingüe--Congresos	1

Imagen 3: Pantalla de resultados de búsqueda con opción "Que empiece con" (Búsqueda alfabética)

- Revisar por código: permite buscar signaturas topográficas dentro de una determinada biblioteca. Como respuesta, el sistema ofrece un listado alfabético de signaturas topográficas. El segundo ítem de la lista es el más coincidente con la signatura buscada. Cada signatura mostrada lleva a su correspondiente registro bibliográfico.

CAPITULO 3: METODOLOGÍA DE ANÁLISIS

Como se mencionó en el primer capítulo, el análisis de bitácoras de transacciones como metodología de análisis en investigaciones, especialmente en el área de la bibliotecología y ciencia de la información, aparece a mediados de los años setenta y desde entonces se han llevado a cabo numerosos estudios para saber cómo se usan los catálogos en línea, sin que los usuarios tengan conocimiento que están siendo observados. Solo acudiendo a la información que se registra en las bitácoras de transacciones de búsqueda estructurada de tal manera que permite la agrupación de datos según preferencias de uso de índices, cadenas de texto usadas, tipos de errores, número de aciertos, etc. es posible determinar patrones de comportamiento con relación a las búsquedas y poder plantear mejoras en los distintos aspectos que tienen impacto en las respuestas que ofrece un catálogo en línea de biblioteca. A continuación se describe la metodología usada en el presente trabajo de investigación.

3.1. Acerca de la metodología

Al tratarse de un análisis de transacciones de bitácoras, se adoptará un enfoque cuantitativo para el análisis de los datos. Los datos a analizar son generados por el software del catálogo en línea de la PUCP.

En función de las búsquedas sin resultados y de la información que pueden brindar las bitácoras en este caso, se va a analizar lo siguiente:

- Frecuencia de búsquedas sin resultados.
- Frecuencia del número de términos usados y su relación con el porcentaje de búsquedas sin resultados.

- Frecuencia de índices usados y su relación con el porcentaje de búsquedas sin resultados.
- Frecuencia de tipos de búsqueda usados y su relación con el porcentaje de búsquedas sin resultados.

A partir de los indicadores mencionados, se trataría de explicar las siguientes situaciones:

- Determinación de las causas de las búsquedas sin resultados durante las semanas de mayor actividad de búsquedas en el catálogo el año 2012.
- Causas de las búsquedas sin resultados y su relación con los índices usados en las búsquedas.

3.2. Las bitácoras de transacciones de búsqueda de e-Library®

Las bitácoras de transacciones que ofrece e-Library básicamente permiten conocer el número de registros que recuperan las búsquedas, asociado a la cadena de texto usado y al índice escogido por el usuario. Además arroja información que se detallará más adelante en el presente capítulo: se explicarán las partes que la conforman con ayuda de muestras, tanto del formato original de la bitácora como el formato final con el que ha sido posible realizar el análisis para el presente trabajo de investigación.

3.2.1. Descripción de los elementos

Como se mencionó en el capítulo 1, la bitácora de transacciones de búsquedas es un archivo de texto generado por el propio sistema informático donde se registra una acción del usuario (palabra o frase buscada) y una respuesta por parte del sistema, que en este caso es el número de registros recuperados. Hay datos sobre otros detalles, gracias a los cuales se ha podido obtener valiosa información de análisis para el presente estudio.

Las bitácoras de transacciones de búsquedas en el catálogo en línea de la PUCP registran la siguiente información:

Fecha y hora: Indica la fecha, hora y los segundos en que se ejecutó el inicio de sesión en el catálogo en línea, y la búsqueda.

Login: Indica si el inicio de sesión en el catálogo en línea fue exitoso.

User: Cuando se ingresa al catálogo sin identificación, en la bitácora se registra el valor “WEBSERVER”, que es el usuario por defecto. Cuando un usuario se registra, aparece el valor correspondiente a su ID de usuario. Todas las sesiones necesitan de un usuario para iniciarse, tal como se muestra en la imagen 4:

```
2012/12/1 00:00:02 Login: Succeeded User: 00213195 Library: EEGGCC Session: 248080064
2012/12/1 00:00:07 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 106730065
```

Imagen 4: Muestra de bitácora de transacciones de inicio de sesión

La imagen 5 muestra otro ejemplo de bitácora, donde se observa la siguiente información:

Library: Indica la biblioteca a la que pertenece el usuario que inició la sesión. En la gran mayoría aparece CENTRAL, que es la biblioteca asociada al usuario WEBSERVER.

Session: Es un número aleatorio de nueve dígitos que el sistema asigna apenas iniciada una sesión. Ese número aparece en todas las transacciones de búsquedas también.

UFSQuery: Su presencia en la bitácora indica el tipo de búsqueda usado. En este caso quiere decir búsqueda por palabra clave o frase. Aquí aparece el nombre del índice usado en la búsqueda. Cuando este nombre no aparece, quiere decir que la búsqueda se realizó desde un formulario de búsqueda externo, como el que está en la página Web principal del Sistema de Bibliotecas de la PUCP. Aunque no aparezca este valor, se considera que el índice empleado es el índice GENERAL, o todos los campos.

```

2012/12/1 01:05:48 UFSQuery: "Sandweiss, Daniel H."Session: 178660028 Server: home Records: 11
2012/12/1 01:05:54 Login: SucceededUser: WEBSEVER Library: CENTRAL Session: 294500030
2012/12/1 01:05:54 UFSQuery: "nunca mas" Session: 294500030 Server: lhome Records: 13
2012/12/1 01:06:24 Login: SucceededUser: WEBSEVER Library: CENTRAL Session: 205530022
2012/12/1 01:06:24 UFSQuery: "antigona furiosa" Session: 205530022 Server: lhome Records: 1
2012/12/1 01:06:38 Login: SucceededUser: WEBSEVER Library: CENTRAL Session: 203610032
2012/12/1 01:06:41 Login: SucceededUser: WEBSEVER Library: CENTRAL Session: 270170061
2012/12/1 01:06:41 UFSQuery: "introducion al derecho" Session: 270170061 Server: lhome Records: 0
  
```

Imagen 5: Muestra de bitácora de transacciones con información adicional (caso 1)

Otro caso se observa en la imagen 6, donde se observa la siguiente información:

ATSQUERY: Su presencia en la bitácora indica que el tipo de búsqueda usado es la búsqueda alfabética.

Qualifier: Solo aparece para la búsqueda alfabética para indicar el índice usado.

Server: En todas las transacciones aparece el valor "lhome". No tiene mayor relevancia.

Records: Indica el número de registros recuperados en la transacción de búsqueda.

Aparece el valor cero si la búsqueda no recuperó nada. Para las búsquedas alfabéticas se omite este valor en la bitácora porque este tipo de búsqueda genera como resultado una lista de títulos, autores, temas, etc. en forma de enlaces, dependiendo del índice escogido, a partir de la cual el usuario escoge la más apropiada y la cual llevará a un conjunto de registros recuperados. Siempre habrá resultados a partir de estos listados.

```

2012/12/1 01:02:38 ATSQuery: "Estrategias para el desarrollo de negocios" Qualifier: TITLE Session:
43050011 Server: lhome
2012/12/1 01:21:28 ATSQuery: "un lugar sin limites internet, democracia y desarrollo" Qualifier: TITLE
Session: 150510017 Server: lhome
2012/12/1 01:22:42 ATSQuery: "justicia, democracia y derechos humanos" Qualifier: SUBJECT Session: 206580063
Server: lhome
2012/12/1 01:22:53 ATSQuery: "justicia, democracia y derechos humanos" Qualifier: TITLE Session: 206580063
Server: lhome|
  
```

Imagen 6: Muestra de bitácora de transacciones con información adicional (caso 2)

3.2.2. Muestra de datos originales

Para la presente investigación solo interesa obtener las líneas correspondientes a búsquedas por palabra clave o frase, donde se indica el número de registros recuperados. Es evidente la necesidad de depurar el archivo de la bitácora para poder obtener solo los datos relevantes para la investigación. La siguiente muestra

corresponde a un minuto de transacciones en la bitácora para el día 1 de diciembre de 2012, 12:20 pm. (Imagen 7)

```

2012/12/1 12:20:00 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 81910011
2012/12/1 12:20:00 UFSQuery: "nacer sabiendo" Session: 81910011 Server: lhome Records: 2
2012/12/1 12:20:04 UFSQuery: "GENERAL Cabe Aristoteles" Session: 256990160 Server: lhome Records: 1
2012/12/1 12:20:05 UFSQuery: "GENERAL PROGRAMAS DE BIENESTAR SOCIAL EN EL SECTRO MINERO" Session:
79080076 Server:lhome Records: 0
2012/12/1 12:19:50 UFSQuery: "GENERAL Metafisica de Aristoteles" Session: 224340152 Server: lhome
Records: 11
2012/12/1 12:20:05 UFSQuery: "GENERAL Metafisica de Aristoteles" Session: 224340152 Server: lhome
Records: 11
2012/12/1 12:20:05 UFSQuery: "GENERAL Metafisica de Aristoteles" Session: 224340152 Server: lhome
Records: 11
2012/12/1 12:20:06 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 65560015
2012/12/1 12:20:06 UFSQuery: "ignacio luis triollo" Session: 65560015 Server: lhome Records: 0
2012/12/1 12:20:07 UFSQuery: "TI microeconomia AND AU samuelson AND AU nordhaus" Session: 304640014
Server: lhome Records: 2
2012/12/1 12:20:08 UFSQuery: "GENERAL estrategias de posicionamiento" Session: 259970131 Server: lhome
Records: 4
2012/12/1 12:20:09 Login: Succeeded User: MOVIL Library: CENTRAL Session: 251140175
2012/12/1 12:20:09 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 85660176
2012/12/1 12:20:09 UFSQuery: "economia kurgman" Session: 85660176 Server: lhome Records: 0
2012/12/1 12:20:16 UFSQuery: "GENERAL resistencia de materiales" Session: 184850032 Server: lhome
Records: 172
2012/12/1 12:20:23 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 150900100
2012/12/1 12:20:23 UFSQuery: "cecilia tovar +" Session: 150900100 Server: lhome Records: 23
2012/12/1 12:20:29 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 248630068
2012/12/1 12:20:29 UFSQuery: "economia kurgman" Session: 248630068 Server: lhome Records: 12
2012/12/1 12:20:30 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 10650131
2012/12/1 12:20:32 UFSQuery: "GENERAL ESTUDIOS CULTURALES" Session: 147790198 Server: lhome Records: 72
2012/12/1 12:20:33 UFSQuery: "GENERAL estudios culturales" Session: 62740023 Server: lhome Records: 72
2012/12/1 12:20:35 UFSQuery: "AU pablo augusto van thienen" Session: 65560015 Server: lhome Records: 0
2012/12/1 12:20:36 UFSQuery: "AU emilio vasquez" Session: 192660110 Server: lhome Records: 28
2012/12/1 12:20:45 UFSQuery: "GENERAL manual del soldador" Session: 10650131 Server: lhome Records: 4
2012/12/1 12:20:45 UFSQuery: "GENERAL portocarrero" Session: 62740023 Server: lhome Records: 459
2012/12/1 12:20:54 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 175610178
2012/12/1 12:20:54 UFSQuery: "cultura y sociedad en la musica popular peruana" Session: 175610178 Server:
lhome Records:1
2012/12/1 12:20:54 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 152020200
2012/12/1 12:20:54 UFSQuery: "el hombre y su lenguaje" Session: 152020200 Server: lhome Records: 3
2012/12/1 12:20:55 Login: Succeeded User: WEBSERVER Library: CENTRAL Session: 7120207
2012/12/1 12:20:55 UFSQuery: "krugman" Session: 7120207 Server: lhome Records: 71
2012/12/1 12:20:56 UFSQuery: "GENERAL facebook" Session: 302760170 Server: lhome Records: 14
2012/12/1 12:20:59 UFSQuery: "GENERAL la republica de platon" Session: 224340152 Server: lhome Records: 6
  
```

Imagen 7: Muestra de bitácora correspondiente a un minuto de transacciones

3.2.3. Muestra de datos procesados

Con ayuda de un script hecho en lenguaje de programación PERL, los datos de las bitácoras adquieren un nuevo formato, fácilmente exportable a una hoja de cálculo (Imagen 8). El proceso de formateo solo deja las líneas correspondientes a las transacciones de búsquedas, omitiéndose las líneas correspondientes a los inicios de sesión. Posteriormente a este proceso, se ha omitido el valor "lhome" de todas las líneas porque es irrelevante.

```

2012/12/1 12:20:00|81910011|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|nacer sabiendo|2|
2012/12/1 12:20:04|256990160|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|Cabe Aristoteles|1|
2012/12/1 12:20:05|79080076|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|PROGRAMAS DE BIENESTAR SOCIAL EN EL SECTRO MINERO|0|
2012/12/1 12:20:05|224340152|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|Metafisica de Aristoteles|11|
2012/12/1 12:20:05|224340152|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|Metafisica de Aristoteles|11|
2012/12/1 12:20:06|65560015|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|ignacio luis triollo|0|
2012/12/1 12:20:07|304640014|CENTRAL|WEBSERVER|UFSQUERY|TI,AU,AU|microeconomia AND AU samuelson AND AU nordhaus|2|
2012/12/1 12:20:08|259970131|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|estrategias de posicionamiento|4|
2012/12/1 12:20:09|85660176|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|economia kurgman|0|
2012/12/1 12:20:16|184850032|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|resistencia de materiales|172|
2012/12/1 12:20:23|150900100|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|cecilia tovar +|23|
2012/12/1 12:20:29|248630068|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|economia krugman|12|
2012/12/1 12:20:32|147790198|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|ESTUDIOS CULTURALES|72|
2012/12/1 12:20:33|62740023|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|estudios culturales|72|
2012/12/1 12:20:35|65560015|CENTRAL|WEBSERVER|UFSQUERY|AU|pablo augusto van thienen|0|
2012/12/1 12:20:36|192660110|CENTRAL|WEBSERVER|UFSQUERY|AU|emilio vasquez|28|
2012/12/1 12:20:45|10650131|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|manual del soldador|4|
2012/12/1 12:20:45|62740023|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|portocarrero|459|
2012/12/1 12:20:54|175610178|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|cultura y sociedad en la musica popular peruana|1|
2012/12/1 12:20:54|152020200|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|el hombre y su lenguaje|3|
2012/12/1 12:20:55|7120207|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|krugman|71|
2012/12/1 12:20:56|302760170|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|facebook|14|
2012/12/1 12:20:59|224340152|CENTRAL|WEBSERVER|UFSQUERY|GENERAL|la republica de platon|6|
    
```

Imagen 8: Muestra de bitácora de transacciones con formato para exportar a hoja de cálculo

La imagen 9 muestra la tabla de datos exportada a M.S. Excel, lista para ser analizada.

FECHA	ID SESION	BIBLIOTECA	ID USUARIO	TIPO BUSQ	INDICE	TERMINOS	ACIERTOS
2012/12/1 12:20:00	81910011	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	nacer sabiendo	2
2012/12/1 12:20:04	256990160	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	Cabe Aristoteles	1
2012/12/1 12:20:05	224340152	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	Metafisica de Aristoteles	11
2012/12/1 12:20:05	79080076	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	PROGRAMAS DE BIENESTAR SOCIAL EN EL 0	0
2012/12/1 12:20:05	224340152	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	Metafisica de Aristoteles	11
2012/12/1 12:20:06	65560015	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	ignacio luis triollo	0
2012/12/1 12:20:07	304640014	CENTRAL	WEBSERVER	UFSQUERY	TI,AU,AU	microeconomia AND AU samuelson AND 2	2
2012/12/1 12:20:08	259970131	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	estrategias de posicionamiento	4
2012/12/1 12:20:09	85660176	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	economia kurgman	0
2012/12/1 12:20:16	184850032	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	resistencia de materiales	172
2012/12/1 12:20:23	150900100	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	cecilia tovar +	23
2012/12/1 12:20:29	248630068	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	economia krugman	12
2012/12/1 12:20:32	147790198	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	ESTUDIOS CULTURALES	72
2012/12/1 12:20:33	62740023	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	estudios culturales	72
2012/12/1 12:20:35	65560015	CENTRAL	WEBSERVER	UFSQUERY	AU	pablo augusto van thienen	0
2012/12/1 12:20:36	192660110	CENTRAL	WEBSERVER	UFSQUERY	AU	emilio vasquez	28
2012/12/1 12:20:45	10650131	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	manual del soldador	4
2012/12/1 12:20:45	62740023	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	portocarrero	459
2012/12/1 12:20:54	175610178	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	cultura y sociedad en la musica popular p	1
2012/12/1 12:20:54	152020200	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	el hombre y su lenguaje	3
2012/12/1 12:20:55	7120207	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	krugman	71
2012/12/1 12:20:56	302760170	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	facebook	14
2012/12/1 12:20:59	224340152	CENTRAL	WEBSERVER	UFSQUERY	GENERAL	la republica de platon	6

Imagen 9: Muestra de la bitácora de transacciones exportadas a M.S. Excel

3.3. Recolección de los datos

La recolección es un proceso automático que va añadiendo los datos a un archivo de bitácoras en tiempo real. Por esta razón, en este caso, el proceso no es

invasivo, y permite obtener datos confiables ya que los usuarios no se sienten observados y actúan de manera natural al momento de hacer sus búsquedas, sin presiones que pudieran afectar los resultados o inducir a errores.

Para la presente tesis se ha recolectado las bitácoras correspondientes a todas las transacciones de búsquedas del año 2012. Estas están consolidadas en archivos de texto comprimidos correspondientes a cada mes del año que en conjunto llegan a tener un tamaño de 79.2 MB. Al ser descomprimidos llegan a 409 MB, lo cual corresponde a 4'459,386 líneas de texto.

3.4. Herramientas de análisis

Para poder hacer el análisis de los datos ha sido necesario darle un formato más legible a las bitácoras de transacciones de búsquedas, las cuales son archivos de texto que tienen una estructura poco legible según se ha visto en las muestras del presente capítulo. Estas necesitan un tratamiento especial con herramientas UNIX y procesadores de texto. Para ello, se usó el script, desarrollado en lenguaje de programación PERL disponible en la wiki de usuarios de SirsiDynix, que da formato a las bitácoras de búsquedas de los sistemas Symphony® o Unicorn®. Para poder analizar los datos se ha hecho uso de esta herramienta (Ver Anexo 15) la que permite especialmente separar los campos por barras, en un formato más limpio y listo para la exportación a hojas de cálculo.

Los archivos procesados con el mencionado script requieren un trabajo de normalización de ciertos valores antes de ser procesados en las hojas de cálculo. Por ejemplo, las búsquedas hechas a través del formulario incrustado en la página Web del Sistema de Bibliotecas de la PUCP apuntan al catálogo bibliográfico en línea y registran las transacciones de búsqueda en las bitácoras pero no se indica el índice usado en la

transacción. La configuración de ese formulario establece que el índice por defecto al cual se dirigen las búsquedas es el llamado *General*. Entonces, en las bitácoras se tiene que llenar ese campo con la palabra GENERAL, como lo tienen el resto de líneas que registran búsquedas en ese índice. Es preferible hacer la normalización en los archivos de texto que en las hojas de cálculo porque el consumo de recursos de procesamiento informático es menor.

Por la gran cantidad de datos con los que se debe trabajar es recomendable usar procesadores de texto especiales, con gran capacidad de procesamiento. En este caso se han usado los programas *UltraEdit* y de *Notepad++*, este último no requiere el pago de licencia para su uso.

La herramienta más importante para el análisis de los datos es la hoja de cálculo. En este caso se ha usado Microsoft Excel 2010. Al respecto, es importante mencionar que la versión 2003 y las anteriores no pueden procesar más de 65,536 filas. Es decir, con esa versión no se puede analizar bitácoras que contienen más líneas que ese número. Tanto la versión 2007 como la versión 2010 pueden procesar hasta 1'048,576 de líneas o filas de texto.

Para analizar los datos se procedió a elaborar varias tablas, las cuales se pasan a describir a continuación:

- Para cada mes del año: transacciones de búsquedas que incluyen los siguientes campos: fecha, biblioteca, ID de usuario, Tipo de búsqueda, Términos de búsqueda, Índice, Aciertos.
- De la tabla anterior se elaboró una similar pero de aquellas transacciones con cero aciertos.

A partir de las tablas anteriores se crearon otras tablas mediante la aplicación de filtros para poder realizar conteos de acuerdo con el uso de fórmulas.

CAPITULO 4: RESULTADOS DEL ANÁLISIS

Después de haber identificado y detallado las partes que conforman la bitácora de transacciones de búsqueda objeto del presente estudio, y haber conseguido un formato de datos más legible, de fácil exportación a hojas de cálculo, el trabajo de análisis se hace más fácil, guiado por el marco teórico y las experiencias encontradas en estudios realizados en otras instituciones. Se presenta a continuación los resultados del análisis realizado.

4.1. Número total de búsquedas

Como se indicó anteriormente, el catálogo permite ejecutar búsquedas por palabras clave, las cuales son de dos tipos: 1) simple y 2) avanzada. En la búsqueda simple existe la opción de búsqueda por listado alfabético (Imagen 3), rotulada como “Que empiece con”, la cual también se registra en las bitácoras de transacciones aunque nunca se indica el número de registros recuperados porque esta opción muestra listas alfabéticas de términos o frases coincidentes (o lo más aproximado posible) con los usados en la búsqueda, pero no recupera registros bibliográficos directamente. Las bitácoras generan líneas con información de búsquedas por palabra clave, en cualquier índice, y lo mismo para las búsquedas alfabéticas. La suma de ambas dan el total de búsquedas realizadas en el catálogo en línea. Durante el 2012, se registraron en total 2'866,649 búsquedas, donde el 99.71% fue por palabra clave; y el 0.29%, por listado alfabético (Tabla 1).

Desde el año 2002, cuando se empezó a medir la frecuencia del número de sesiones y búsquedas en el catálogo en línea objeto de este estudio, se observó que mayo y setiembre eran los meses con la mayor cantidad de sesiones y búsquedas

durante el año (siendo setiembre el de mayor actividad). Esto coincide con el periodo de exámenes parciales y finales en los semestres académicos 1 (marzo-julio) y 2 (agosto-diciembre) respectivamente (Gráfico 1). Las horas de mayor actividad son entre las 9 a.m. y las 2 p.m., llegando a concentrar el 45.35% de todas las sesiones para el 2012 (Gráfico 2).

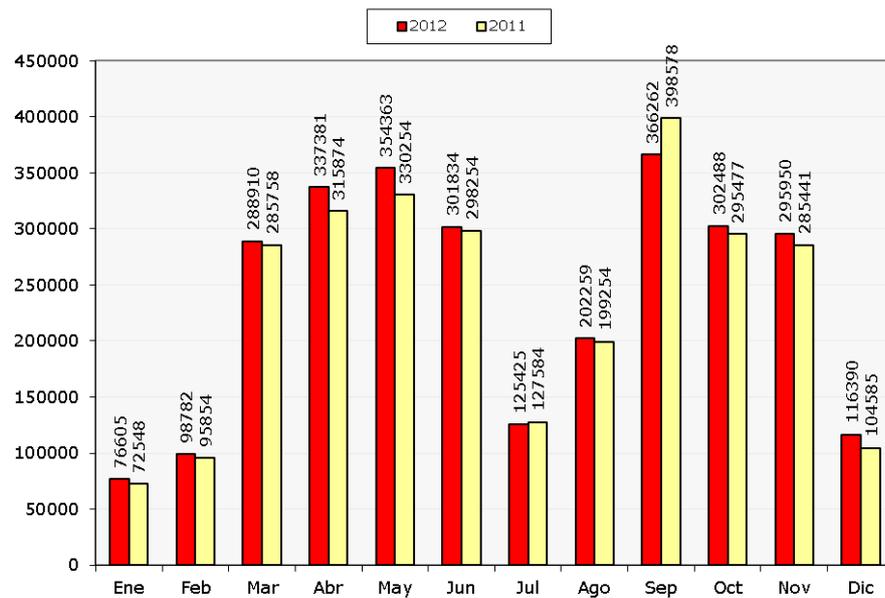


Gráfico 1: Comparativo mensual de número de búsquedas para los años 2011 y 2012

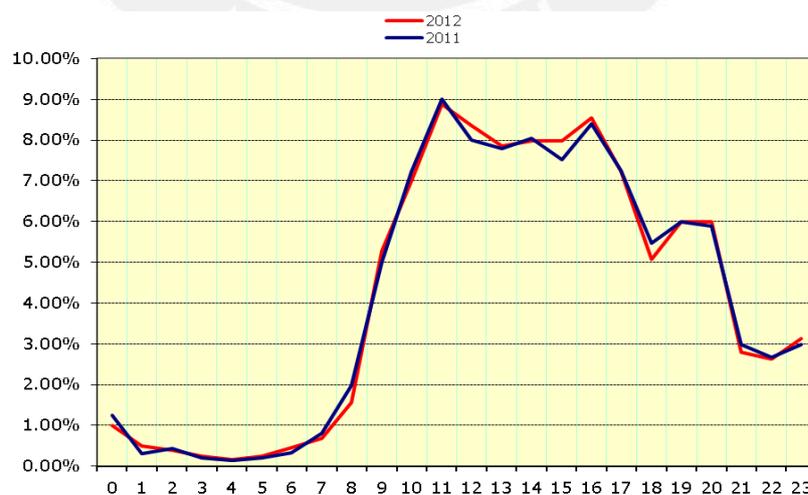


Gráfico 2: Comparativo de sesiones por horas del día en e-Library para los años 2011 y 2012

El análisis de los datos para la presente investigación se ha hecho solo tomando en cuenta el número de búsquedas por palabra o frase porque las bitácoras solo registran el número de aciertos recuperados para estas búsquedas.

	Total de búsquedas	%
Palabra o frase	2'858,377	99.71%
Lista alfabética	8,272	0.29%
Total	2'866,649	

Tabla 1: Número de búsquedas totales para el año 2012

4.2. Proporción de búsquedas con resultados vacíos

Cuando se tomó una muestra de datos el año 2009 para saber el porcentaje de búsquedas con resultados vacíos, el análisis básico entonces aplicado arrojó que poco más del 30% de transacciones de búsquedas no ofrecían resultados a los usuarios. Para el año 2012, la cifra había disminuido unos pocos puntos porcentuales a 26.46%. Es decir, de un total de 2'858,377 de búsquedas realizadas, 756,221 tuvieron cero aciertos, tal como se aprecia en la siguiente tabla:

	Total
# de búsquedas por palabra clave	2'858,377
# de búsq. resultados vacíos	756,221
Cero aciertos %	26.46%

Tabla 2: Proporción de resultados vacíos en búsquedas por palabra clave el año 2012

En el siguiente gráfico se muestra el número de búsquedas a lo largo del año y su relación respecto a aquellas búsquedas que no arrojaron resultados. El detalle con cifras mensuales se puede ver en el Anexo 1.

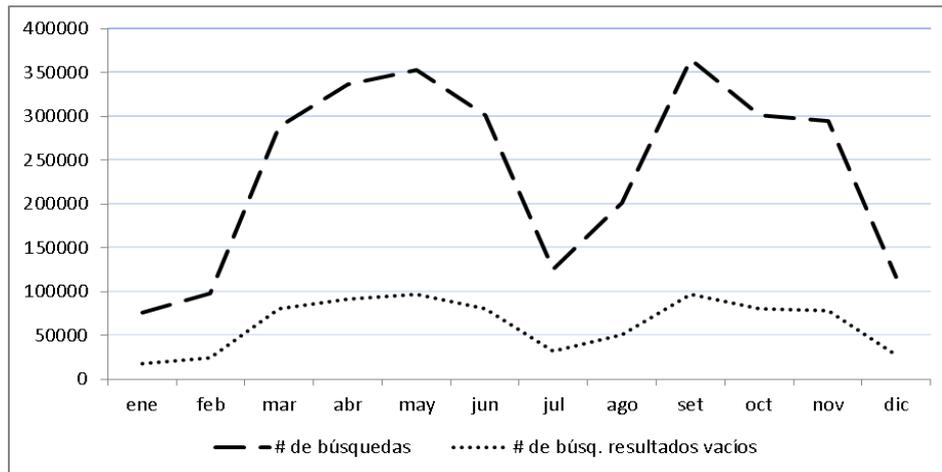


Gráfico 3: Comparación mensual del número de búsquedas frente al número de búsquedas con resultados vacíos

El promedio anual de la proporción de búsquedas con resultados vacíos es poco más de un cuarto, específicamente 26.46% como se mencionó anteriormente. Durante el año estudiado el rango va de 23.65% a 27.91%. Se puede decir que de cada cuatro búsquedas, una no obtiene resultados. Como puede apreciarse en el Gráfico 4, a pesar de las variaciones en cantidad mes a mes, la proporción de búsquedas con resultados vacíos se mantiene pareja a lo largo del año analizado (2012).

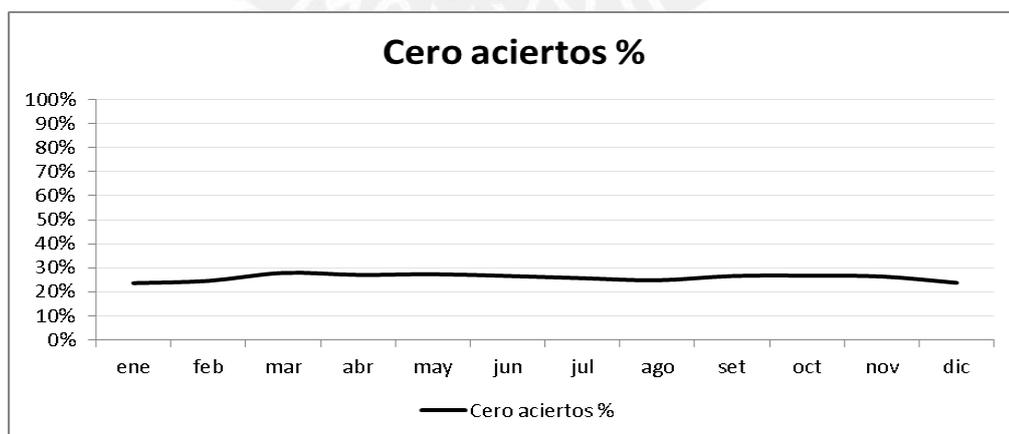


Gráfico 4: Proporción mensual de búsquedas sin resultados

Coincidentemente, los meses con menor número de búsquedas presentan la menor proporción de resultados vacíos: enero (23.65%) y diciembre (23.77%). Sin embargo, no sucede lo mismo con los meses que tienen mayor número de búsquedas. La proporción más alta la obtuvo el mes de marzo con el 27.91%, pero los meses de mayo y setiembre, que son los meses con mayor número de búsquedas, registraron 27.35% y 26.60% respectivamente. Para mayores detalles ver el Anexo 2.

4.3. Número de búsquedas por tipo (simple, avanzada)

La mayor cantidad de búsquedas se registra en el tipo de búsqueda simple con 98.73% por ser el tipo de búsqueda por defecto. El uso de la búsqueda avanzada es muy bajo pero es el que tiene mayor proporción de resultados vacíos como se ve a continuación:

Tipo de búsqueda	# de búsquedas	%
Simple	2'822,178	98.73%
Avanzada	36,199	1.27%
Total	2'858,377	100%

Tabla 3: Número de búsquedas por tipo

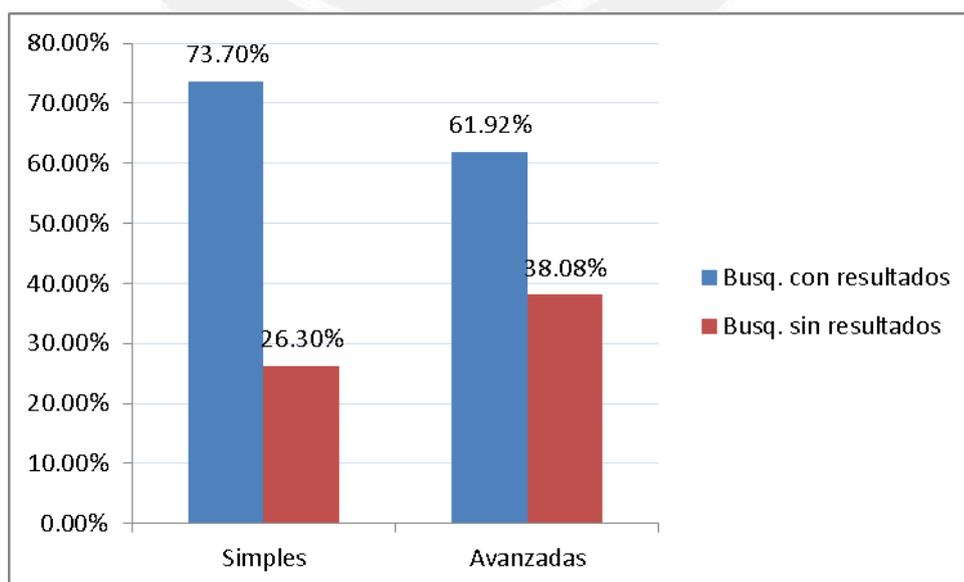


Gráfico 5: Proporción de búsquedas con resultados vacíos por tipo de búsqueda

Solo el 1.27% de las búsquedas corresponde a la búsqueda avanzada. El poco uso de esta opción puede explicar en parte la elevada proporción de resultados vacíos que ésta consigue: 38.08% frente al 26.30% de la búsqueda simple, la cual es muy similar a la proporción anual de ambos tipos de búsqueda juntos.

4.4. Búsquedas por índice (general, autor, título, tema, revista-periódico, serie, editorial)

El índice más usado para las búsquedas es el *general* o etiquetado como *todos los campos* con el 87.29%. Se podría decir que de diez búsquedas realizadas, casi nueve se hacen usando este índice. No se puede hablar de una preferencia en este caso porque éste es el índice configurado por defecto para el tipo de búsqueda más usado (búsqueda simple), también por defecto. La proporción de búsquedas sin resultados para este índice durante el 2012 es del 25.48%. Esta cifra representa el 84.06% del total de resultados vacíos. Los detalles para cada mes se pueden ver en el Anexo 3.

El índice con mayor preferencia, pero el segundo más usado, es el de *autor*. Su uso llega al 5.63% del total de búsquedas realizadas, con una proporción de 24.41% de resultados vacíos, más de dos puntos porcentuales por debajo de la proporción anual de todas las búsquedas sin resultados juntas. Los detalles para cada mes se pueden ver en el Anexo 4.

El índice de *título* es el tercero más usado, con una proporción de 3.60% y con 35.06% de búsquedas sin resultados. Los detalles para cada mes se pueden ver en el Anexo 5.

El índice de *tema* es el cuarto más usado, con una proporción de 1.50% y con 49.62% de búsquedas sin resultados. Aquí se podría decir que cinco de cada diez

búsquedas por tema no muestran resultados. Los detalles para cada mes se pueden ver en el Anexo 6.

La combinación de índices, la que se realiza en el formulario de búsqueda avanzada (Imagen 2), es la quinta más usada con 1.27% del total de las búsquedas realizadas y con 38.08% de resultados vacíos. Los detalles para cada mes se pueden ver en el Anexo 7.

El uso del índice *revista-periódico* es el sexto más usado con 0.61% del total de las búsquedas realizadas, y con 45.97% de resultados vacíos. Los detalles para cada mes se pueden ver en el Anexo 8.

El uso del índice *serie* es muy bajo, con 0.09% del total de búsquedas realizadas pero tiene la más alta proporción de recuperación vacía con 82.19%. Los detalles para cada mes se pueden ver en el Anexo 9.

Con 0.02% del total de las búsquedas realizadas, el índice *editorial* es el menos usado. Tiene una recuperación vacía 10.36%, la proporción más baja de todas. Para los detalles, ver el Anexo 10. Es pertinente mencionar que este índice se encuentra en la búsqueda avanzada pero nunca fue combinado con otros índices. Esto podría significar que quienes lo usaron tienen mayor nivel de experiencia o conocen mejor el catálogo.

El 15.94% de las búsquedas con resultados vacíos pertenecen a las realizadas en los índices de *autor*, *título*, *tema*, *combinados*, *revista-periódico*, *serie* y *editorial* juntos.

Llama la atención el bajo uso de los índices de *autor*, *título* y *tema*. Juntos representan el 10.73% de las búsquedas realizadas. También es bajo su uso en comparación con otros estudios. En el de Hildreth, el uso del índice de *título* llegó a 30.5%, casi igual al uso de palabra clave o índice general con 30.6%, aunque la cantidad de búsquedas analizadas fueron solo 467 (1997: 60). En el estudio de Villen-Rueda,

Senso y de Moya (2007: 334) el uso del índice de *título* llegó a 49%, por delante del de *autor* con 37% y del de *tema* con 14%. Antell y Huang solo estudiaron las búsquedas por *tema*, encontrando que solo se usaron el 4.6% de las veces con 48.8% de resultados vacíos (2008: 71).

Índice	No. de Búsquedas	% de Búsquedas	Con resultados		Sin resultados	
			Cantidad	%	Cantidad	%
General	2'494,948	87.29%	1'859,410	74.53%	635,682	25.48%
Autor	160,818	5.63%	121,570	75.59%	39,248	24.41%
Título	103,005	3.60%	66,892	64.94%	36,113	35.06%
Tema	42,832	1.50%	21,578	50.38%	21,254	49.62%
Combinados	36,266	1.27%	22,456	61.92%	13,810	38.08%
Revista-Periódico	17,497	0.61%	9,453	54.03%	8,044	45.97%
Serie	2,448	0.09%	436	17.81%	2,012	82.19%
Editorial	563	0.02%	505	89.70%	58	10.30%
Total	2'858,377	100.00%	2'102,300	73.55%	756,221	26.46%

Tabla 4: Proporción de resultados vacíos de las búsquedas en todos los índices en relación a la frecuencia de su uso para el periodo de enero a diciembre del 2012.

En el siguiente gráfico se presenta la relación entre la frecuencia de uso de los índices y el porcentaje de recuperación vacía. Con la excepción del uso combinado de índices y el índice de *editorial*, se observa que a menor frecuencia de uso de los índices, mayor es el porcentaje de recuperación vacía.

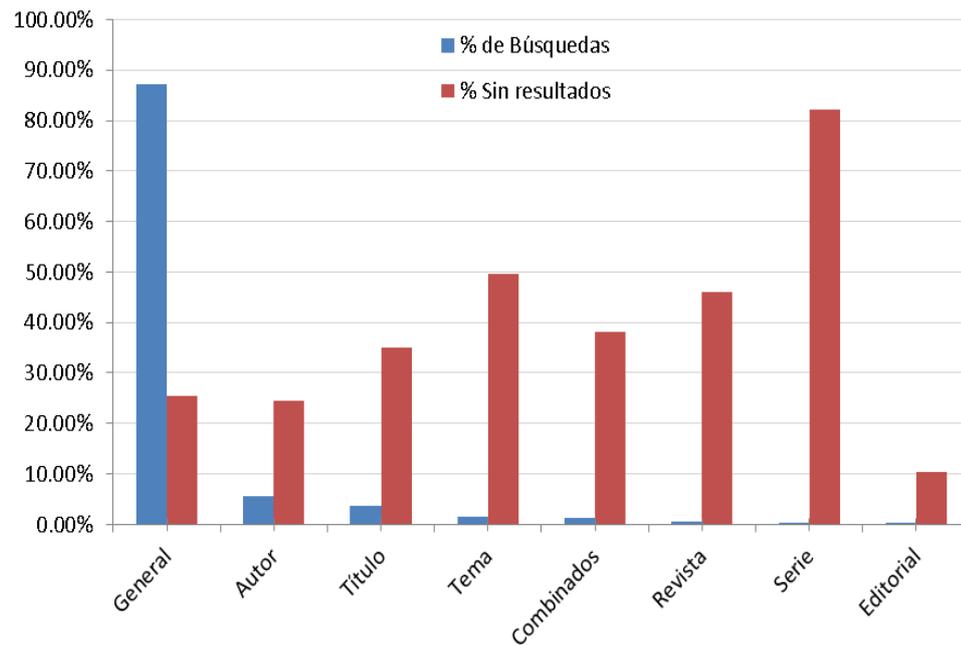


Gráfico 6: Relación entre la frecuencia de uso de índices y el porcentaje de recuperación vacía para el periodo de enero a diciembre de 2012.

Este análisis necesita ser complementado con el que determina las causas de la recuperación vacía. Se trata de un análisis donde solo se ha tomado muestras de las semanas con mayor número de búsquedas, durante los meses de mayor actividad.

4.5. Proporción de resultados vacíos en las búsquedas por el número de palabras usadas (longitud de las consultas)

La relación entre el número de palabras y el porcentaje de resultados vacíos resulta muy interesante de ser analizado en la presente investigación, aunque no se ha determinado que el número de palabras usado en las búsquedas tenga efecto en el número de resultados vacíos.

Hildreth hizo un análisis del número de palabras usadas en las búsquedas con sus respectivos porcentajes de resultados vacíos (1997: 61). Las búsquedas con dos palabras se usaron en la mayoría de los casos (41.3%) pero fueron las búsquedas con cinco palabras las que tuvieron mayor porcentaje de resultados vacíos (100%). Se debe indicar

que solo se analizaron 467 búsquedas para las cifras señaladas líneas arriba. En el estudio de Lau y Goh la búsqueda con dos palabras fue la más usada (34.05%) y también fue la que obtuvo mayor porcentaje de resultados vacíos (14.37%) (2006: 1321). Moulaison muestra que en las búsquedas de su estudio mayormente se usaron de una a tres palabras, aunque no se establece el porcentaje de resultados vacíos para cada una de ellas (2008: 235).

Al igual que en el estudio de Lau y Goh (2006), en la presente investigación se encontró que las búsquedas donde se usaron dos palabras generaron el mayor porcentaje de resultados vacíos, aunque con distinto porcentaje, 28.56%. Cuando se usaron tres palabras, el porcentaje de resultados vacíos llegó al 20.24%. A partir del uso de dos palabras en las búsquedas, el porcentaje de resultados vacíos va disminuyendo (Tabla 5)

A diferencia del estudio de Hildreth (1997), aquí se encontró que las búsquedas con cinco palabras generaron el 9.58% de los resultados vacíos. En el 12.33% de las búsquedas con resultados vacíos se usó una palabra.

El máximo número de palabras usadas en una búsqueda fue 53, en el mes de setiembre solo una vez.

Meses	Número de palabras							
	1	2	3	4	5	6	7	7+
Todos	12.33%	28.56%	20.24%	14.00%	9.58%	5.64%	3.35%	6.31%

Tabla 4: Proporción de resultados vacíos en las búsquedas según el número de palabras usadas (enero-diciembre 2012)

En el Anexo 11 se puede ver el detalle para cada mes, los cuales tienen mucha similitud con los porcentajes del consolidado anual.

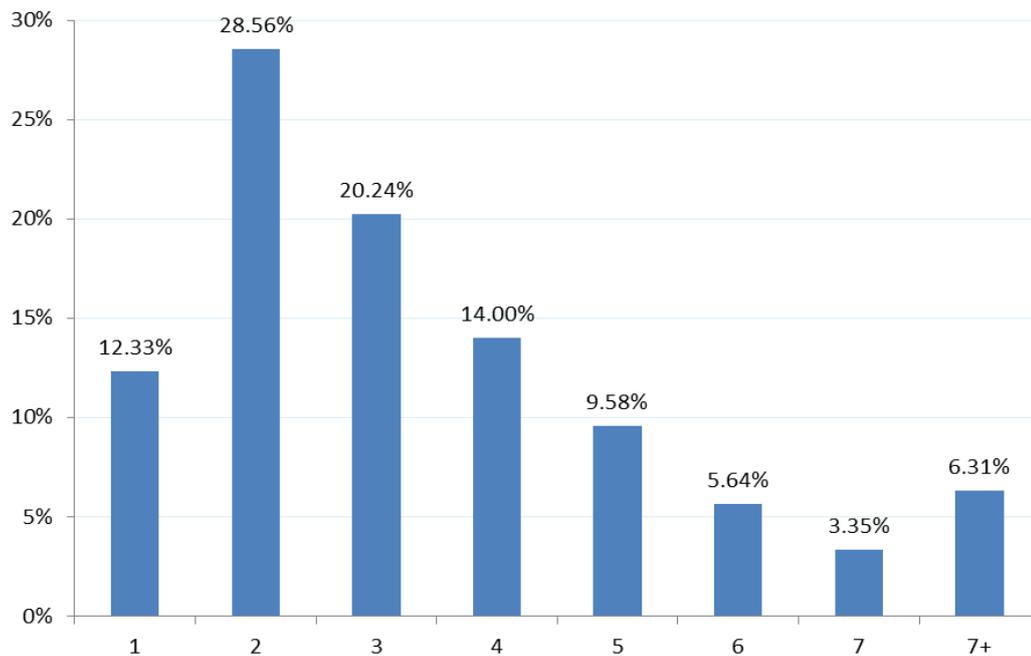


Gráfico 7: Proporción de resultados vacíos en las búsquedas según el número de palabras usadas (enero-diciembre 2012)

4.6. Identificación de causas que generaron recuperación vacía en las búsquedas

Para identificar las causas que generaron recuperación vacía en las búsquedas se optó por analizar las bitácoras correspondientes a las semanas que registraron mayor cantidad de búsquedas de los meses con mayor número de búsquedas durante el 2012, es decir, mayo y setiembre.

Debido a que este análisis no se hace mediante fórmulas sino mediante una revisión visual, y a la gran cantidad de datos registrados para los periodos que se mencionó, se escogió para el análisis solo las transacciones de búsquedas realizadas durante las horas de mayor actividad, que son, para ambos meses, de 9 a.m. a 2 p.m. Al mes de mayo le corresponden 5,946 transacciones de búsquedas o líneas analizadas,

para los días del 21 al 25. A setiembre, 6,575 transacciones de búsquedas los días del 24 al 28.

Tras el análisis se llegó a identificar tres causas que generaron recuperación vacía en las búsquedas:

1. Términos No Coincidentes: aquí se está considerando todas aquellas palabras o frases escritas correctamente en las fórmulas de búsqueda que no coinciden con las usadas en la descripción bibliográfica y terminología temática asignada a cada registro. Es decir, son casos donde el usuario formula bien su búsqueda, emplea palabras adecuadas y pertinentes a los índices seleccionados, pero a pesar de ello, la recuperación es vacía. El uso incorrecto de mayúsculas, minúsculas, u omisión de tildes y/o signos de puntuación no se ha tomado en cuenta porque el motor de búsqueda los pasa por alto. Por ejemplo, éste recuperará la misma cantidad de registros si se busca la palabra *Perú* o *peru*. También pasa por alto la sintaxis: el orden de las palabras no altera el resultado.

Para poder determinar que el nombre y/o apellidos de un autor estaban bien escritos se procedió a buscarlos en la Web para la confirmación respectiva. El mismo trabajo se hizo para cada uno de los otros índices.

2. Escritura incorrecta: se refiere a los errores tipográficos y ortográficos, omitiendo las excepciones indicadas en el punto anterior. La mayoría de casos señalados como escritura incorrecta corresponde a errores tipográficos: orden de las letras en las palabras, presencia o ausencia de espacios entre letras o palabras, según el caso, introducción de caracteres extraños, etc.

3. Elección del índice incorrecto: aquí se refiere a la búsqueda de palabras o frases en índices no pertinentes. Por ejemplo, se ha encontrado casos de búsqueda de

nombres de personas en el índice de títulos de publicaciones periódicas, o de temas en el índice de autores.

4.6.1 Causas de búsquedas sin resultados según índices usados durante las semanas con mayor número de búsquedas

4.6.1.1. Causas de búsquedas sin resultados por uso del índice *tema*

Llama mucho la atención que la búsqueda por *tema* tenga poco más del 50% de recuperación vacía en el análisis global. Si se obvia la búsqueda por *serie*, que es un caso muy particular y se comentará más adelante, *tema* es el índice con el más alto porcentaje de recuperación vacía. En mayo y setiembre, el porcentaje de resultados vacíos para las búsquedas fue de 54.26% y 50.52% respectivamente (Anexo 6).

El análisis para el índice *tema* de las muestras de los meses de mayo y setiembre, arrojó que el 19.52% y 21.95% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto del índice) mientras que el 80.48% (mayo) y el 78.05% (setiembre) de recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los términos usados para la descripción temática de los registros (Tabla 6).

Con este resultado es inevitable pensar en el grado de efectividad del lenguaje controlado de este sistema para la descripción de temas. En el análisis de la terminología se encontró un uso muy frecuente de lenguaje natural. La siguiente lista es una muestra de los términos usados en las búsquedas temáticas y que no obtuvieron resultados:

- motivacion y emoción
- la constitución del 93
- informalidad en el peru

- liderazgo que obtiene resultados
- snacks de frutas deshidratadas
- análisis de los poemas de cesar vallejo
- etapas de desarrollo del niño

Se puede apreciar la influencia que tienen los motores de búsqueda Web en las búsquedas realizadas en el catálogo en línea estudiado. El análisis de Kumar en un grupo de tres universidades de la India reportó que del 54.4% al 62.4% es el rango de quienes declararon no saber que el catálogo en línea funciona de manera distinta a un motor de búsqueda Web (2011: 17-18). En una de las universidades estudiadas, el 77.6% buscó información en el catálogo en línea como lo hacen en los motores de búsqueda como Google. En las otras dos universidades este porcentaje llegó a poco más del 48%. Esto demostraría que las búsquedas en los motores de búsqueda Web tienen una fuerte influencia en el usuario al momento de usar un catálogo en línea. En otro estudio llevado a cabo solo para la biblioteca de la Universidad de Punjabi, India, y siguiendo la misma metodología, Kumar llega a las mismas conclusiones, con cifras muy similares (2012: 56-70). Años antes se hizo un estudio sobre los problemas en las búsquedas por tema en un catálogo en línea y su relación con los motores de búsqueda Web en la Universidad Estatal de California, Los Ángeles. También se concluyó que el modelo que traen en mente los usuarios al usar el catálogo en línea está influenciado por la experiencia previa de los motores de búsqueda Web, y que en la mayoría de los casos los usuarios no son conscientes de estas diferencias, razón por la cual tienen las mismas expectativas y conductas en el uso de ambos modelos (Yu y Yang 2004: 178). Los resultados obtenidos en la presente investigación confirmarían que la misma situación se presenta en la PUCP.

Si bien es cierto que el total de búsquedas temáticas representa el 1.50% del total, no quiere decir que no se busquen temas al hacer uso del índice general.

4.6.1.2. Causas de recuperación vacía para las búsquedas en el índice *título*

En el 2012 las búsquedas por *título* tuvieron un porcentaje de recuperación vacía del 35.06% en el análisis global. En mayo y setiembre, el porcentaje de resultados vacíos fue de 29.11% y 29.62% respectivamente (Anexo 5). Del total de búsquedas sin resultados por *título*, de las muestras de los meses de mayo y setiembre, arrojó que el 21.11% y 17.39% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto del índice). El 78.90% (mayo) y el 82.61% (setiembre) de recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los títulos de los registros bibliográficos (Tabla 6).

4.6.1.3. Causas de recuperación vacía para las búsquedas por combinación de índices

Como se muestra en la Tabla 4, las búsquedas por combinación de índices tienen un porcentaje anual de recuperación vacía de 38.08%. En mayo y setiembre, el porcentaje de resultados vacíos es de 39.08% y 41.89% respectivamente (Anexo 7). El análisis para las búsquedas por combinación de índices, de las muestras de los meses de mayo y setiembre, arrojó que el 21.28% y 23.14% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto de índice). El 78.72% (mayo) y el 76.87% (setiembre) de recuperación vacía se debió a que los términos empleados en la búsqueda no coincidieron con los términos empleados en la descripción de los registros bibliográficos (Tabla 6).

En la muestra de mayo se encontró que los usuarios del catálogo efectuaron 18 combinaciones distintas 94 veces. En el 100% de los casos se usó el operador booleano AND. La combinación *Autor AND Título* produjo el 32.98% del total de los resultados

vacíos para la combinación de índices, seguida de *General AND Autor* (14.89%) y *Título AND Tema* (10.64%).

Para la muestra de setiembre se realizaron 27 combinaciones distintas 135 veces. En el 100% de los casos se usó el operador booleano AND. La combinación *Autor AND Título* produjo el 34.81% del total de los resultados vacíos para la combinación de índices, seguida de *General AND Autor* (14.07%) y *Tema AND Tema* (10.37%).

4.6.1.4. Causas de recuperación vacía para las búsquedas en el índice *general*

Las búsquedas en el índice *general* tuvieron un porcentaje anual de recuperación vacía de 25.48%. En mayo y setiembre, el porcentaje de resultados vacíos fue de 27.29% y 26.34% respectivamente (Anexo 3). El análisis para las búsquedas en el índice *general* de las muestras de los meses de mayo y setiembre, arrojó que el 24.93% y 20.22% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto de índice). El 75.07% (mayo) y el 79.78% (setiembre) de recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los términos empleados en la descripción de los registros bibliográficos (Tabla 6).

4.6.1.5 Causas de recuperación vacía para las búsquedas en el índice de *autor*

Las búsquedas en el índice de *autor* tuvieron un porcentaje anual de recuperación vacía del 24.41%. En mayo y setiembre, los porcentajes fueron de 24.04% y 24.79% respectivamente (Anexo 4). El análisis para el índice de *autor*, de las muestras de los meses de mayo y setiembre, arrojó que el 71.31% y 59.44% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto del índice). Se observó que hay muchas dificultades para escribir nombres y apellidos en otros idiomas

distintos al español. El 28.69% (mayo) y el 40.56% (setiembre) de recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los nombres de los autores en la descripción de los registros bibliográficos (Tabla 6).

4.6.1.6 Causas de recuperación vacía para las búsquedas en el índice de *revista-periódico*

Las búsquedas en el índice de *revista-periódico* tuvieron un porcentaje anual de recuperación vacía del 45.97%. En mayo y setiembre, el porcentaje de resultados vacíos fue de 47.98% y 46.99% respectivamente (Anexo 8). El análisis para el índice *revista-periódico*, de las muestras de mayo y setiembre, arrojó que el 85.46% y 72.23% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto del índice). Se observó que, aparte de los títulos de las publicaciones periódicas, se busca junto a ellos el número y/o volumen, los cuales no están indizados. Estos resultados sugerirían que los usuarios no comprenden bien su uso en gran parte porque no se especifica en la etiqueta del formulario de búsqueda que éste solo sirve para buscar títulos de publicaciones periódicas. El 14.55% (mayo) y el 27.78% (setiembre) de la recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los títulos de las publicaciones periódicas en los registros bibliográficos (Tabla 6).

4.6.1.7 Causas de recuperación vacía para las búsquedas en el índice de *serie*

Las búsquedas en el índice *serie* tuvieron un porcentaje anual de recuperación vacía del 82.19%. En mayo y setiembre, el porcentaje de resultados vacíos fue de 87.16% y 86.01% respectivamente (Anexo 9). El análisis para el índice de *serie* para las muestras de los meses de mayo y setiembre, el 100% y 97.62% respectivamente se debió a errores de los usuarios (escritura incorrecta y uso incorrecto de índice). Se

observó que en este índice se buscan nombres de series de televisión, lo que sugeriría que no se comprende bien para qué sirve este índice. El 2.38% (setiembre) de la recuperación vacía se debió a que los términos empleados en las búsquedas no coincidieron con los títulos de las series en los registros bibliográficos (Tabla 6).

Índice	Meses	Términos No Coincidentes	Escritura incorrecta	Índice incorrecto
Tema	Mayo	80.48%	11.59%	7.93%
	Setiembre	78.05%	12.80%	9.15%
Título	Mayo	78.90%	16.06%	5.05%
	Setiembre	82.61%	12.61%	4.78%
Combinadas	Mayo	78.72%	19.15%	2.13%
	Setiembre	76.87%	20.90%	2.24%
General	Mayo	75.07%	23.53%	1.40%
	Setiembre	79.78%	19.11%	1.11%
Autor	Mayo	28.69%	44.62%	26.69%
	Setiembre	40.56%	31.82%	27.62%
Revista-Periódico	Mayo	14.55%	41.82%	43.64%
	Setiembre	27.78%	46.30%	25.93%
Serie	Mayo	0.00%	7.14%	92.86%
	Setiembre	2.38%	7.14%	90.48%

Tabla 5: Consolidado de las causas de recuperación vacía en las búsquedas por índices usados los meses de mayo y setiembre de 2012, durante las semanas de mayor cantidad de búsquedas (mayo, 21 al 25; setiembre, 24 al 28).

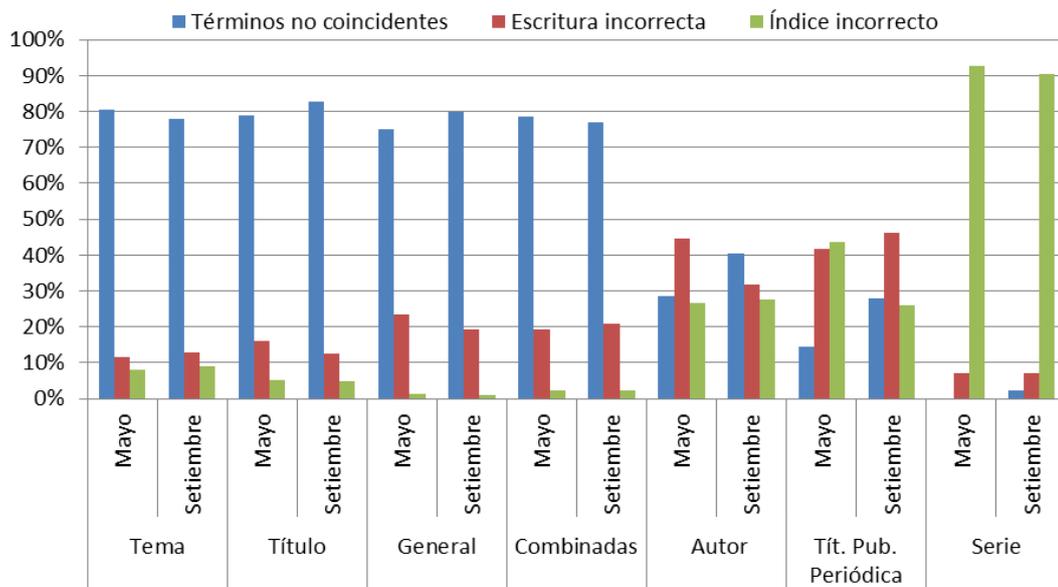


Gráfico 8: Consolidado de las causas de recuperación vacía en las búsquedas por índices usados los meses de mayo y setiembre de 2012, durante las semanas de mayor cantidad de búsquedas

En las muestras analizadas no se encontraron búsquedas realizadas en el índice *editorial*.

CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

5.1.1. Sobre el análisis de bitácoras de transacciones como metodología

El análisis de bitácoras de transacciones es una metodología que se aproxima a los cincuenta años de uso, aunque en el ámbito de la bibliotecología y ciencia de la información, específicamente en catálogos en línea, recién empezó a aplicarse a fines de los años setenta. A partir de entonces se han realizado muchos estudios de catálogos en línea, principalmente en universidades de Norteamérica, Europa y Asia. Para el caso peruano, no se ha podido encontrar estudios usando esta metodología.

El uso de esta metodología presenta ciertas dificultades, principalmente porque las bitácoras no tienen un formato fácil de exportar a hojas de cálculo. Se requiere un trabajo previo que implica el uso de técnicas, de nivel intermedio y avanzado, en elaboración de *scripts* para obtener el formato adecuado para su análisis. Existen tantos formatos de bitácoras de transacciones como software de catálogos en línea hay en el mercado. No existe un estándar que establezca un formato único para estas bitácoras de manera tal que se pueda crear software capaz de analizarlas sin importar el fabricante del software de los catálogos en línea.

La recolección de datos de la interacción entre los usuarios y el catálogo se da dentro de un proceso automático y no invasivo, el cual es una ventaja en sí mismo porque el usuario no se siente observado y, por consiguiente, actúa de una manera natural. Esta metodología no permite saber detalles de las circunstancias bajo las cuales se lleva a cabo una o más búsquedas por el mismo usuario en un tiempo determinado, a diferencia de la observación directa en el mismo instante en que se realizan las

búsquedas y donde la recolección de información puede complementarse con entrevistas.

No es posible saber, mediante el análisis de las bitácoras de transacciones de búsquedas, el grado de relevancia de las respuestas dadas por el sistema. Las búsquedas que sí obtienen resultados no necesariamente satisfacen las necesidades de información del usuario. Los resultados de las búsquedas, muchos o pocos, podrían bien ser considerados por el usuario como resultados vacíos también, porque no responden a sus necesidades.

Los estudios de uso de catálogos en línea con esta metodología son muy variados porque es posible aplicarlos a varios objetos de estudio, pero también dependen de los datos que las bitácoras pueden ofrecer. Los más comunes están orientados a determinar las preferencias y errores de los usuarios durante sus búsquedas, problemas de uso del catálogo, patrones de uso y las causas de búsquedas no exitosas, entre otros.

5.1.2. Sobre la interacción del usuario y el catálogo en línea del Sistema de Bibliotecas de la PUCP

El catálogo en línea del Sistema de Bibliotecas de la PUCP tiene un motor de búsquedas bastante potente, basado en BRS/Search, que permite usar operadores booleanos, de posición (de palabras) y relacionales (para números) con opción al truncamiento y sustitución de caracteres, tipificados en búsqueda simple y búsqueda avanzada.

El usuario del catálogo del Sistema de Bibliotecas de la PUCP aprovecha mínimamente el potencial que ofrece el motor de búsqueda. Hay un predominio mayoritario del tipo de búsqueda simple, con una frecuencia de uso del 98.73% sobre la

búsqueda avanzada. Del mismo modo, la búsqueda por palabra clave predomina con 99.71% del total realizado sobre la búsqueda alfabética. Este predominio se puede explicar porque tanto el tipo de búsqueda básica como la búsqueda por palabra clave están predeterminados en el formulario de búsqueda de la pantalla inicial del catálogo.

El uso de operadores booleanos es muy escaso con apenas con el 1.27% del total de búsquedas realizadas por palabra clave, a pesar de que el formulario de búsqueda avanzada incluye estos operadores para que el usuario pueda combinar términos de búsqueda. Asimismo, el uso de operadores de posición y relacionales, así como las opciones de truncamiento y sustitución es prácticamente nulo.

5.1.3. Sobre las búsquedas sin resultados

Se ha identificado tres acciones en las búsquedas que causan recuperación vacía: el uso de términos no coincidentes, la escritura incorrecta de los términos de búsqueda, y el uso incorrecto de índices. La escritura incorrecta de los términos de búsqueda y el uso incorrecto de índices constituyen propiamente errores cometidos por los usuarios.

La proporción de 26.46% de búsquedas sin resultados, del total de búsquedas, puede considerarse moderada en comparación con las cifras obtenidas en otros estudios similares. A pesar de que el número de búsquedas realizadas varía significativamente mes a mes, la proporción de búsquedas sin resultados varía poco, fluctuando entre 23.65% y 27.91% a lo largo del año.

La búsqueda avanzada presenta una proporción de resultados vacíos alta, con 38.08%. Sin embargo, su escaso uso no es determinante en el resultado global de este indicador.

La preferencia de uso de los índices *autor*, *título* y *tema* va en ese orden y es muy baja. Juntos llegan al 10.73% del total de búsquedas realizadas. Con respecto a

estos tres índices, mientras más baja es la frecuencia de uso, más alta es la proporción de búsquedas sin resultados.

Respecto a la cantidad de palabras usadas para formular una búsqueda, se encontró que a partir de dos palabras, a mayor número de palabras usadas en las búsquedas, menor es la proporción de resultados vacíos. Es posible que el uso de varias palabras en las fórmulas de búsqueda indique que los usuarios están buscando registros que saben de antemano que existen en el catálogo.

Para las semanas más activas de los meses de mayo y setiembre, el 53.17% de las causas de recuperación vacía corresponde al uso de términos no coincidentes. El 46.83% corresponde a errores cometidos por los usuarios (uso de índice incorrecto y errores de escritura). Es decir, los errores de los usuarios no son la causa principal de las búsquedas sin resultados en el catálogo en línea de la PUCP. En estos periodos, la búsqueda por los índices de *tema*, *título*, *general* y combinados tuvieron una proporción de recuperación vacía por encima del 75% a causa del uso de términos de búsqueda no coincidentes con los usados en la descripción bibliográfica. En cambio, las búsquedas por los índices de *autor*, *revista-periódico* y *serie* tuvieron una proporción de recuperación vacía entre el 59% y 100% a causa de errores cometidos por los usuarios.

La proporción de búsquedas sin resultados para el índice *tema* llegó a poco más del 50%, del cual casi el 80% se debió a que los términos usados por los usuarios en las búsquedas no coincidieron con los términos usados en la descripción temática de los registros bibliográficos por parte de la biblioteca. Esto pone en evidencia que los usuarios no están familiarizados con los encabezamientos de materia empleados en la indización de los registros bibliográficos, y más bien están usando mucho el lenguaje natural, tal como lo hacen en los motores de búsqueda como Google, donde siempre reciben respuestas aunque no necesariamente relevantes en primera instancia.

5.2 Recomendaciones

Tomando en cuenta los resultados obtenidos del análisis de las bitácoras de búsquedas del catálogo de la PUCP, y la determinación de las causas de las búsquedas sin resultados, a continuación se da un conjunto de recomendaciones con el objetivo de disminuir la proporción de resultados vacíos en este catálogo, y de sugerir otros estudios que ayuden a mejorar el catálogo y las capacitaciones sobre su uso, entre otros.

5.2.1. Capacitaciones de uso del catálogo en línea

Si bien es cierto que el Sistema de Bibliotecas de la PUCP brinda capacitaciones en uso del catálogo a los alumnos ingresantes y a quienes lo solicitan, es importante que se ponga énfasis en ejecutar las siguientes recomendaciones:

- Explicar el funcionamiento y ventajas de las búsquedas mediante selección de índices tanto a nivel de búsqueda simple como a nivel de búsqueda avanzada. Aunque en algunos casos parece obvio para los bibliotecarios, los resultados han demostrado que no se comprende bien su uso, especialmente para el caso de los índices de *autor*, *revista-periódico* y *serie*.
- Profundizar en la enseñanza del funcionamiento de la búsqueda avanzada y el uso de los operadores booleanos que el catálogo ofrece.
- Explicar la lógica detrás de los encabezamientos de materia y cómo sacar provecho de su característica hipertextual en la ficha electrónica.

5.2.2. Mejoras en la interfaz de búsqueda

Durante la realización de este estudio, y a la luz de los resultados del análisis, se ve necesario sugerir las siguientes mejoras en la interfaz de búsqueda a fin de reducir los errores que puedan cometer los usuarios durante la interacción con el catálogo:

- Mejorar los nombres de los rótulos de los índices *serie* y *revista-periódico*. *serie bibliográfica* y *título de revista o periódico* pueden ayudar al usuario a entender mejor para qué sirven estos índices.
- Asignar enlaces a ayudas que expliquen el funcionamiento de cada uno de los elementos de la interfaz de búsqueda a manera de pequeñas ventanas de diálogo, amigables y superpuestas al formulario de búsqueda para evitar otras ventanas en el navegador.
- Ver la posibilidad de reemplazar la actual interfaz por una que utilice lógica difusa (*fuzzy logic*) para la corrección automática de errores ortográficos y tipográficos, con la consiguiente reducción del número de resultados vacíos en las búsquedas. Enterprise y Portfolio son las soluciones existentes como reemplazo de e-Library® para los clientes de SirsiDynix Corp.

5.2.3. Activar el módulo de control de autoridades temáticas en el catálogo

La elevada proporción de búsquedas sin resultados mediante el índice *tema* pone de manifiesto la necesidad de implementar, a la brevedad posible, el módulo de control de autoridades temáticas en el catálogo. Esto sería de gran ayuda para el usuario porque de esta manera recibiría sugerencias por parte del catálogo para replantear la fórmula de búsqueda usada que no recuperó registros bibliográficos, a la vez que le permite familiarizarse con la terminología usada por los catalogadores, reduciendo el uso de términos no coincidentes en las formulaciones de sus búsquedas, una de las principales causas de la obtención de resultados vacíos. Mientras tanto, se debe recomendar a los usuarios el uso de la búsqueda alfabética (*browse*) para que se familiaricen con el lenguaje empleado en los encabezamientos de materia.

A manera de complemento, y en vista de los resultados obtenidos del análisis para la presente investigación, se recomienda también el uso de lenguaje no controlado en la indexación temática de los registros bibliográficos. Para saber qué palabras o frases están usando los usuarios del catálogo y que no obtienen resultados. Esta información la proporciona la misma bitácora de transacciones de búsqueda del catálogo.

5.2.4. Estudios mensuales de términos no coincidentes para mejorar el desarrollo de colecciones

Para cada índice usado es posible determinar qué se está buscando y no se encuentra en el catálogo debido al uso de términos no coincidentes entre lo que se busca y lo que existe en el catálogo. El análisis de las bitácoras de búsquedas puede ayudar a determinar, por ejemplo, qué nombres de autores y qué títulos se buscan con más frecuencia y no se encuentran en el catálogo. Esto constituye una fuente de información valiosa para detectar las debilidades de la colección bibliográfica, una herramienta de mucha ayuda para el desarrollo de colecciones.

5.2.5. Estudios complementarios a la presente investigación

Para la presente investigación se cumplió el objetivo de poder determinar las causas de las búsquedas con resultados vacíos en el catálogo en línea del Sistema de Bibliotecas de la PUCP. Sin embargo, debido a que el potencial que ofrece el análisis de las bitácoras de transacciones de búsquedas del catálogo en línea, objeto de este estudio, es bastante amplio, se sugiere determinar patrones de comportamiento de los usuarios con respecto a las búsquedas, que estén relacionados a las distintas características que éstos tienen, tales como perfil, facultad, especialidad, biblioteca a la que pertenece (del

campus o fuera de él), etc. y poder detectar qué fortalezas y debilidades tienen los usuarios cuando usan el catálogo para buscar información.

Para el presente estudio no se pudo hacer este tipo de análisis porque la gran mayoría de búsquedas son anónimas y no se obliga a los usuarios a identificarse para poder usar el catálogo. Como se vio en la descripción de las bitácoras de transacciones, un dato importante es el ID de usuario que éstas capturan, el cual es suficiente para poder cruzar información de la bitácoras con la base de datos de usuarios donde está la información de perfil, facultad, especialidad, etc. Con fines de estudio, se puede configurar el catálogo para que los usuarios se identifiquen obligatoriamente antes de realizar búsquedas, dejando la opción anónima a los visitantes. De esta manera se podría responder interrogantes como ¿qué tan efectivas son las capacitaciones en búsquedas en el catálogo brindadas a los alumnos ingresantes a la universidad?, ¿los alumnos de qué facultades tienen menor o mayor proporción de búsquedas sin resultados?, ¿existe relación entre el nivel académico de los usuarios y el desempeño de sus búsquedas en el catálogo?, ¿en qué grado mejoran los usuarios su desempeño en las búsquedas en el catálogo a lo largo de su vida universitaria?, etc.

5.3 Comentario final

Siempre las bibliotecas se han preocupado por saber cómo sus usuarios buscan información en sus catálogos y si sus interfaces de búsqueda son adecuadas y amigables para ellos. Tradicionalmente se recurre a las entrevistas y encuestas para realizar diagnósticos de uso y desempeño de los usuarios frente a los catálogos en línea. Sin embargo, por lo trabajoso y demandante que son estas herramientas, su aplicación es escasa.

En los últimos años, las bibliotecas de la región y específicamente del país, han optado por contar con sistemas integrados de gestión de bibliotecas con capacidad para registrar bitácoras de transacciones de búsquedas, entre otras cosas. Varias de ellas tienen la materia prima almacenada en sus servidores esperando su recolección para ser analizada, con gran potencial para realizar diagnósticos básicos o más profundos que van a permitir plantear mejoras para sus capacitaciones de usuarios, para sus interfaces de búsqueda y para sus propias colecciones. En contextos donde los bibliotecarios tienen más responsabilidades y servicios que atender, esta metodología, por su carácter automático y no invasivo durante la etapa de recolección de datos, ofrece ventajas sobre las entrevistas y encuestas, aunque estas últimas pueden complementar estudios con información que las bitácoras no registran.

Como no se ha encontrado estudios publicados basados en esta metodología a nivel regional y de país, se espera que el presente trabajo de investigación anime a las bibliotecas a realizar trabajos similares, especialmente en aquellas que usan la misma interfaz de búsqueda que tiene el Sistema de Bibliotecas de la PUCP, siempre apuntando a mejorar uno de los principales servicios que toda biblioteca tiene: el catálogo en línea y su interfaz de búsqueda.

BIBLIOGRAFÍA

- ANTELL, Karen y Jie HUANG
2008 "Subject Searching Success Transaction Logs, Patron Perceptions, and Implications for Library Instruction". *Reference & User Services Quarterly*. Chicago, 2008, volumen 48, número 1, pp. 68-76.
- BABU, B. Ramesh y Ann O'BRIEN
2000 "Web OPAC interfaces: an overview". *The Electronic Library*. Bingley, 2000, volumen 18, número 5, pp. 316-330.
- BEHESHTI, Jamshid
1997 "The evolving OPAC". *Cataloging & Classification Quarterly*. Abington, 1997, volumen 24, número 1-2, pp. 163-185.
- BORGMAN, Christine L.
1996 "Why Are Online Catalogs Still Hard to Use?" *Journal of the American Society for Information Science*. Hoboken, 1996, volumen 47, número 7, pp. 493-503.
- BORLUND, Pia
2003 "The Concept of Relevance in IR". *Journal of the American Society for Information Science and Technology*. Hoboken, 2003, volumen 54, número 10, pp. 913-925.
- CHURCH, Karen, Barry SMYTH, Paul COTTER y Keith BRADLEY
2007 "Mobile information access: A study of emerging search behavior on the mobile Internet". *ACM Transactions on the Web*. New York, 2007, volumen 1, número 1, pp. 1-38.
- COOPER, William S.
1988 "Getting beyond Boole". *Information Processing & Management*. Amsterdam, 1988, volumen 24, número 3, 243-248.
- DICKSON, Jean
1984 "An Analysis of User Errors in Searching an Online Catalog". *Cataloging & Classification Quarterly*. Abington, 1984, volumen 4, número 3, pp. 19-38.
- HILDRETH, Charles R.
1997 "The use and understanding of keyword searching in a university online catalog". *Information Technology and Libraries*. Chicago, 1997, volumen 16, número 2, pp. 52-63.

KASKE, Neal K.

1993 “Research methodologies and transaction log analysis: issues, questions, and a proposed model”. *Library Hi Tech*. Bingley, 1993, volumen 11, número 2, pp. 79-86.

KERN-SIMIRENKO, Cheryl

1983 “OPAC User Logs: Implications for Bibliographic Instruction”. *Library Hi Tech*. Bingley, 1983, volumen 1, número 3, pp. 27-35.

KUHLTHAU, Carol C.

1993 “A Principle Of Uncertainty For Information Seeking”. *Journal of Documentation*. Bingley, 1993, volumen 49, número 4, pp. 339–355.

1991 “Inside the search process: Information seeking from the user’s perspective”. *Journal of the American Society for Information Science*. Hoboken, 1991, volumen 45, número 2, pp. 361–371.

KUMAR, Shiv

2012 “Impact of internet search engines on OPAC users: a study of Punjabi University, Patiala (India)”. *Program: Electronic Library and Information Systems*. Bingley, 2012, volumen 46, número 1, pp. 56-70.

2011 “Effect of web searching on the OPAC: a comparison of selected university libraries”. *Library Hi Tech New*. Bingley, 2011, volumen 28, número 6, pp. 14–21.

KURTH, Martin

1993 “The limits and limitations of transaction log analysis”. *Library Hi Tech*. Bingley, 1993, volumen 11, número 2, pp. 98-104.

LAU, Eng Pwey y Dion Hoe-Laian GOH

2006 “In search of query patterns: A case study of a university OPAC”. *Information Processing & Management*. Amsterdam, 2006, volumen 42, número 5, pp. 1316–1329.

MALLIARI, Aphrodite y Daphne KYRIAKY-MANESSI

2007 “Users' behaviour patterns in academic libraries' OPACs: a multivariate statistical analysis”. *New Library World*. Bingley, 2007, volumen 108, número 3-4, pp. 107–122.

- MIZZARO, Stefano
1997 “Relevance : The Whole History”. *Journal of the American Society for Information Science*. Hoboken, 1997, volumen 48, número 9, pp. 810-832.
- MOULAISON, Heather L.
2008 “OPAC Queries at a Medium-Sized Academic Library: A Transaction Log Analysis”. *Library Resources & Technical Services*. Chicago, 2008, volumen 52, número 4, pp. 230-237.
- ORTIZ-REPISO, Virginia, Virginia BAZÁN, Agnes PINSATI y Mario COTTEREAU
2006 “How researchers are using the OPAC of the Spanish Council for Scientific Research Library Network”. *The Electronic Library*. Bingley, 2006, volumen 24, número 2, pp. 190-211.
- PETERS, Thomas A.
1993 “The history and development of transaction log analysis”. *Library Hi Tech*. Bingley, 1993, volumen 11, número 2, pp. 41-66.
1989 “When Smart People Fail: An Analysis of the Transaction Log of an Online Public Access Catalog”. *Journal of Academic Librarianship*. Oxford, 1989, volumen 15, número 5, pp. 267-273.
- PETERS, Thomas A., Martin KURTH, Patricia FLAHERTY, Beth SANDORE y Neal K. KASKE
1993 “An introduction to the special section on transaction log analysis”. *Library Hi Tec*. Bingley, 1993, volumen 11, número 2, pp. 38-40.
- PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ. BIBLIOTECA CENTRAL.
2007 “Análisis comparativo de uso de iLink: 2005 y 2006”. *Bibliopucp*. Lima, 2007, número 7-8, pp. 42-47
- SALVADOR OLIVÁN, José A.
2008 *Recuperación de la información*. Primera edición. Buenos Aires: Alfagrama.
- SANDERSON, Mark y W. Bruce CROFT
2012 “The History of Information Retrieval Research”. *Proceedings of the IEEE*. New York, 2012, número 100, pp. 1444-1451.
- SANDORE, Beth
1993 “Applying the results of transaction log analysis”. *Library Hi Tec*. Bingley, 1993, volumen 11, número 2, pp. 87-97.

- SANDORE, Beth, Patricia FLAHERTY, Neal K. KASKE, Martin KURTH y Thomas A. PETERS
1993 “A manifesto regarding the future of transaction log analysis”. *Library Hi Tech. Bingley*, 1993, volumen 11, número 2, pp. 105-107.
- SARACEVIC, Tefko
2012 “Research on relevance in information science: A historical perspective”. En T. Carbo, & T. Bellardo Hahn (Ed.), *Proceedings of the American Society for Information Science and Technology (ASIS&T) 2012 Pre-conference on the History of ASIS&T and Information Science and Technology*. pp. 49-60.
- VILLEN-RUEDA, Luis, José A. SENSO y Félix DE MOYA-ANEGÓN
2007 “The Use of OPAC in a Large Academic Library: A Transactional Log Analysis Study of Subject Searching”. *The Journal of Academic Librarianship*. Oxford, 2007, volumen 33, número 3, pp. 327–337.
- WIKIPEDIA CONTRIBUTORS
2012 *Wikipedia, The Free Encyclopedia*. BRS/Search. Consulta: 22 de noviembre de 2012.
<<http://en.wikipedia.org/w/index.php?title=BRS/Search&oldid=538220162>>
- WILSON, T. D.
2010 “Fifty years of information behavior research”. *Bulletin of the American Society for Information Science and Technology*. Hoboken, 2010, volumen 36, número 3, pp. 27-34.
- 1999 “Models in information behaviour research”. *Journal of Documentation*. Bingley, 1999, volumen 55, número 3, pp. 249-270.
- YU, Holly y Margo YOUNG
2004 “The Impact of Web Search Engines on Subject Searching in OPAC”. *Information Technology and Libraries*. Chicago, 2004, volumen 23, pp. 168-180.
- ZINK, S. D.
1991 “Monitoring user search success through transaction log analysis: The Wolfpac example”. *Reference Services Review*. Bingley, 1991, volumen 19, número 1, pp. 49-56.

ANEXO 1:

Proporción mensual de búsquedas con resultados vacíos para el año 2012

Meses	# de búsquedas	# de búsq. resultados vacíos	% resultados vacíos
enero	76,274	18,035	23.65%
febrero	98,366	24,183	24.58%
marzo	288,123	80,414	27.91%
abril	336,480	91,075	27.07%
mayo	353,360	96,655	27.35%
junio	301,061	80,151	26.62%
julio	125,012	32,137	25.71%
agosto	201,674	50,081	24.83%
setiembre	365,236	97,148	26.60%
octubre	301,593	80,677	26.75%
noviembre	295,151	77,934	26.40%
diciembre	116,047	27,587	23.77%
Total	2'858,377	756,077	26.45%

ANEXO 2:

Detalle de la proporción mensual de búsquedas con resultados vacíos

Meses	# de búsquedas	# de búsq. resultados vacíos	% Cero aciertos
enero	76,274	18,035	23.65%
febrero	98,366	24,183	24.58%
marzo	288,123	80,414	27.91%
abril	336,480	91,075	27.07%
mayo	353,360	96,655	27.35%
junio	301,061	80,151	26.62%
julio	125,012	32,137	25.71%
agosto	201,674	50,081	24.83%
setiembre	365,236	97,148	26.60%
octubre	301,593	80,677	26.75%
noviembre	295,151	77,934	26.40%
diciembre	116,047	27,587	23.77%
Total	2'858,377	756,077	26.45%

ANEXO 3:

**Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice
general (2012)**

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	61,908	47,155	76.17%	14,753	23.83%
febrero	79,367	59,858	75.42%	19,509	24.58%
marzo	236,095	176,014	74.55%	60,081	25.45%
abril	286,007	209,151	73.13%	76,856	26.87%
mayo	307,082	223,266	72.71%	83,816	27.29%
junio	262,034	192,935	73.63%	69,099	26.37%
julio	105,130	78,670	74.83%	26,460	25.17%
agosto	166,463	125,243	75.24%	41,220	24.76%
setiembre	317,875	234,155	73.66%	83,720	26.34%
octubre	263,307	193,824	73.61%	69,483	26.39%
noviembre	257,963	190,934	74.02%	67,029	25.98%
diciembre	99,479	76,122	76.52%	23,357	23.48%
Total	2'442,710	1'807,327	73.99%	635,383	26.01%

ANEXO 4:

**Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice
autor (2012)**

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	5,497	4,263	77.55%	1,234	22.45%
febrero	7,549	5,820	77.10%	1,729	22.90%
marzo	19,696	14,952	75.91%	4,744	24.09%
abril	18,889	14,137	74.84%	4,752	25.16%
mayo	17,834	13,546	75.96%	4,288	24.04%
junio	15,577	11,733	75.32%	3,844	24.68%
julio	8,267	6,263	75.76%	2,004	24.24%
agosto	13,744	10,471	76.19%	3,273	23.81%
setiembre	17,701	13,313	75.21%	4,388	24.79%
octubre	14,969	11,164	74.58%	3,805	25.42%
noviembre	14,242	10,580	74.29%	3,662	25.71%
diciembre	6,853	5,328	77.75%	1,525	22.25%
Total	160,818	121,570	75.59%	39,248	24.41%

ANEXO 5:

Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice
título

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	3,201	2,253	70.38%	948	29.62%
febrero	4,646	3,368	72.49%	1,278	27.51%
marzo	13,158	3,368	25.60%	9,790	74.40%
abril	12,502	8,824	70.58%	3,678	29.42%
mayo	10,979	7,783	70.89%	3,196	29.11%
junio	9,665	6,857	70.95%	2,808	29.05%
julio	5,257	3,640	69.24%	1,617	30.76%
agosto	9,031	6,617	73.27%	2,414	26.73%
setiembre	11,597	8,162	70.38%	3,435	29.62%
octubre	9,390	6,524	69.48%	2,866	30.52%
noviembre	9,587	6,627	69.12%	2,960	30.88%
diciembre	3,992	2,869	71.87%	1,123	28.13%
Total	103,005	66,892	64.94%	36,113	35.06%

ANEXO 6:

**Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice
tema (2012)**

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	954	538	56.39%	416	43.61%
febrero	2,261	1,513	66.92%	748	33.08%
marzo	6,013	3,417	56.83%	2,596	43.17%
abril	5,325	2,545	47.79%	2,780	52.21%
mayo	4,825	2,207	45.74%	2,618	54.26%
junio	4,174	1,945	46.60%	2,229	53.40%
julio	1,853	939	50.67%	914	49.33%
agosto	2,966	1,645	55.46%	1,321	44.54%
setiembre	5,083	2,510	49.38%	2,573	50.62%
octubre	3,833	1,713	44.69%	2,120	55.31%
noviembre	4,079	1,876	45.99%	2,203	54.01%
diciembre	1,466	730	49.80%	736	50.20%
Total	42,832	21,578	50.38%	21,254	49.62%

ANEXO 7:

Detalle mensual de la proporción de resultados vacíos en las búsquedas de índices combinados

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	744	455	61.16%	289	38.84%
febrero	1,647	1,210	73.47%	437	26.53%
marzo	5,346	3,402	63.64%	1,944	36.36%
abril	5,063	3,153	62.28%	1,910	37.72%
mayo	4,043	2,463	60.92%	1,580	39.08%
junio	3,261	2,016	61.82%	1,245	38.18%
julio	1,438	914	63.56%	524	36.44%
agosto	3,502	2,308	65.91%	1,194	34.09%
setiembre	3,946	2,293	58.11%	1,653	41.89%
octubre	3,159	1,800	56.98%	1,359	43.02%
noviembre	2,842	1,709	60.13%	1,133	39.87%
diciembre	1,275	733	57.49%	542	42.51%
Total	36,266	22,456	61.92%	13,810	38.08%

ANEXO 8:

Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice revista- periódico (2012)

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	929	603	64.91%	326	35.09%
febrero	903	513	56.81%	390	43.19%
marzo	2,000	1,032	51.60%	968	48.40%
abril	1,715	859	50.09%	856	49.91%
mayo	1,907	992	52.02%	915	47.98%
junio	1,484	752	50.67%	732	49.33%
julio	1,222	696	56.96%	526	43.04%
agosto	1,266	742	58.61%	524	41.39%
setiembre	2,175	1,153	53.01%	1,022	46.99%
octubre	1,697	870	51.27%	827	48.73%
noviembre	1,537	809	52.64%	728	47.36%
diciembre	662	432	65.26%	230	34.74%
Total	17,497	9,453	54.03%	8,044	45.97%

ANEXO 9:

Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice serie (2012)

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	74	20	27.03%	54	72.97%
febrero	99	25	25.25%	74	74.75%
marzo	318	45	14.15%	273	85.85%
abril	277	53	19.13%	224	80.87%
mayo	257	33	12.84%	224	87.16%
junio	219	35	15.98%	184	84.02%
julio	104	30	28.85%	74	71.15%
agosto	158	35	22.15%	123	77.85%
setiembre	386	54	13.99%	332	86.01%
octubre	239	36	15.06%	203	84.94%
noviembre	250	48	19.20%	202	80.80%
diciembre	67	22	32.84%	45	67.16%
Total	2,448	436	17.81%	2,012	82.19%

ANEXO 10:

Detalle mensual de la proporción de resultados vacíos en las búsquedas del índice editorial (2012)

Meses	Búsquedas totales	Con resultados		Sin resultados	
		Cantidad	%	Cantidad	%
enero	21	17	80.95%	4	19.05%
febrero	0	0	0.00%	0	0.00%
marzo	47	45	95.74%	2	4.26%
abril	87	73	83.91%	14	16.09%
mayo	54	43	79.63%	11	20.37%
junio	64	61	95.31%	3	4.69%
julio	57	56	98.25%	1	1.75%
agosto	55	53	96.36%	2	3.64%
setiembre	56	47	83.93%	9	16.07%
octubre	61	55	90.16%	6	9.84%
noviembre	37	32	86.49%	5	13.51%
diciembre	24	23	95.83%	1	4.17%
Total	563	505	89.70%	54	9.59%

ANEXO 11:

Detalle mensual de la proporción de resultados vacíos en las búsquedas según el número de palabras usadas (2012)

Meses	Número de palabras							
	1	2	3	4	5	6	7	7+
Ene	13.09%	29.88%	20.78%	12.76%	8.93%	5.30%	3.30%	5.97%
Feb	13.90%	29.71%	19.90%	13.78%	9.01%	5.14%	3.00%	5.56%
Mar	13.17%	29.23%	19.14%	14.31%	9.63%	5.33%	3.31%	5.89%
Abr	12.08%	28.23%	20.32%	14.17%	9.91%	5.66%	3.35%	6.27%
May	10.70%	27.09%	20.36%	14.72%	10.72%	6.29%	3.66%	6.46%
Jun	11.30%	27.87%	21.11%	14.63%	9.77%	5.61%	3.41%	6.30%
Jul	12.16%	28.55%	20.38%	13.86%	9.31%	5.76%	3.34%	6.64%
Ago	14.12%	28.62%	18.95%	13.65%	9.39%	5.63%	3.19%	6.45%
Set	12.57%	28.40%	19.96%	13.84%	9.47%	5.66%	3.32%	6.78%
Oct	11.73%	27.99%	20.25%	14.08%	9.99%	5.82%	3.60%	6.55%
Nov	10.46%	26.95%	21.90%	14.79%	9.96%	6.04%	3.46%	6.44%
Dic	12.68%	30.21%	19.78%	13.41%	8.86%	5.47%	3.20%	6.39%
Anual	12.33%	28.56%	20.24%	14.00%	9.58%	5.64%	3.35%	6.31%

ANEXO 12:

**Comparativo del número de búsquedas realizadas en el catálogo en línea de la
PUCP para los años 2009 y 2010**

Meses	2009	2010	%
Enero	385,906	21,861	-94.34
Febrero	126,429	115,014	-9.03
Marzo	261,729	270,861	3.49
Abril	414,943	367,587	-11.41
Mayo	384,221	303,065	-21.12
Junio	367,016	296,173	-19.30
Julio	172,923	126,876	-26.63
Agosto	246,756	218,718	-11.36
Setiembre	465,033	392,726	-15.55
Octubre	348,642	283,621	-18.65
Noviembre	325,226	270,918	-16.70
Diciembre	122,386	92810	-24.17
TOTAL	3'621,210	2'760,230	-23.78

ANEXO 13:

**Comparativo del número de búsquedas realizadas en el catálogo en línea de la
PUCP para los años 2010 y 2011**

Meses	2010	2011	%
Enero	21,861	72,548	231.86
Febrero	115,014	95,854	-16.66
Marzo	270,861	285,758	5.50
Abril	303,065	315,874	4.23
Mayo	367,587	330,254	-10.16
Junio	296,173	298,254	0.70
Julio	126,876	127,584	0.56
Agosto	218,718	199,254	-8.90
Setiembre	392,726	398,578	1.49
Octubre	283,621	295,477	4.18
Noviembre	270,918	285,441	5.36
Diciembre	92,810	104,585	12.69
TOTAL	2'760,230	2'809,461	1.78

ANEXO 14:

**Comparativo del número de búsquedas realizadas en el catálogo en línea de la
PUCP para los años 2011 y 2012**

Meses	2011	2012	%
Enero	72,548	76,605	5.59
Febrero	95,854	98,782	3.05
Marzo	285,758	288,910	1.10
Abril	315,874	337,381	6.81
Mayo	330,254	354,363	7.30
Junio	298,254	301,834	1.20
Julio	127,584	125,425	-1.69
Agosto	199,254	202,259	1.51
Setiembre	398,578	366,262	-8.11
Octubre	295,477	302,488	2.37
Noviembre	285,441	295,950	3.68
Diciembre	104,585	116,390	11.29
TOTAL	2'809,461	2'866,649	2.04

ANEXO 15:

**Script que da formato a las bitácoras de transacciones de búsqueda del sistema
 Unicorn o Symphony**

```
#!/usr/bin/perl
# written by: Joel Hahn, Niles Public Library District
# last modified: 3 December 2004

if ($ARGV[0] ne "") {
    print "parseweblogs.pl accepts SIRSI's .weblog files as input and writes the
following as output:\n";
    print "Time stamp|Session #|Library|User|Search type|Search index(es)|Search
string|Server|# of hits|\n";
    print "  Search Type:\n";
    print "    UFSQuery = Keyword search\n";
    print "    ATSQuery = Browse search\n";
    print "  Since browse searches have no number of hits, those lines have 'BROWSE'
as the # of hits\n";
    print "  Since weblog files start at midnight and users might already be logged
in, if\nthe login information is not in
the current file, then the Library and User will\nbe output as '-'\n";
    exit;
}

@lines = <STDIN>;

foreach $line (@lines) {
    @fields = (split(/\t/, $line));
    chomp($fields[8]);
    if ($fields[1] eq "Login:") {
        $session = $fields[8];
        $library{$session} = $fields[6];
        $user{$session} = $fields[4];
    #   print "$session|$fields[6]|$fields[4]|$library{$session}|$user{$session}|\n";
    }
    elsif ($fields[1] eq "UFSQuery:") {
        if ($library{$fields[4]} eq "") { $library{$fields[4]} = "-"; }
        if ($user{$fields[4]} eq "") { $user{$fields[4]} = "-"; }
        #####
        $keyindex = "";
        if ($fields[2] =~ / AND | OR | NOT | XOR /) {
            @searchbits = (split(/ /, $fields[2]));
            for ($j=0; $j<= $#searchbits - 2; $j++) {
                if (($searchbits[$j] =~ /^AND$|^OR$|^NOT$|^XOR$/) && ($searchbits[$j+2] !~
/^AND$|^OR$|^NOT$|^XOR$/)) {
                    if ((index("AU,TI,SU,SER,PER,GENERAL", $searchbits[$j+1]) > -1) ||
($searchbits[$j+1] =~ /^[0-9][0-9][0-9]$/))
                    {
                        $keyindex .= $searchbits[$j+1] . ",";
                    }
                }
            }
        }
        if (index($fields[2], "(" > 0) {
            $ftemp = $fields[2];
            while (index($ftemp, "(" > 0) {
                $pos = index($ftemp, "(";
                $ftemp = substr($ftemp, $pos + 1);
                $pos2 = index($ftemp, ")");
                if ($pos2 > 0) {
                    $keyindex .= "(" . substr($ftemp, 0, $pos2) . ")," ;
                }
            }
        }
        #####
        ($keyindex2,$search) = (split(/ /, $fields[2], 2));
        #if ($keyindex ne "") {
            $search = substr($fields[2], 1);
        #}
    }
}
```

```
$keyindex = substr($keyindex2, 1) . "," . $keyindex;
chop($keyindex);
chop($search);
if ($keyindex eq "\\") {
    $keyindex = "NONE";
    $search = "";
}
print
"$fields[0]|$fields[4]|$library{$fields[4]}|$user{$fields[4]}|UFSQUERY|$keyindex|$search|$fields[6]|$fields[8]|\n";
}
else { # ATSQuery
    $fields[2] = substr($fields[2], 1);
    chop($fields[2]);
    print
"$fields[0]|$fields[6]|$library{$fields[6]}|$user{$fields[6]}|ATSQUERY|$fields[4]|$fields[2]|$fields[8]|BROWSE|\n";
}
}
```

