

# Evaluación de colecciones en una biblioteca universitaria utilizando la minería de datos

**Carlos Antonio Pacheco Gómez\***  
**Luis Fernando Morales Mendoza\*\***  
**Paulina Martínez Isidro\*\***  
**Rene Lopez Flores \*\***

*Artículo recibido:*  
20 de noviembre de 2018

*Artículo aceptado:*  
4 de marzo de 2019

*Artículo de investigación*

## RESUMEN

El desarrollo de colecciones es un proceso que permite a las bibliotecas contar con el material bibliográfico adecuado para atender las necesidades de su comunidad de usuarios. En este proceso, una de las restricciones principales es la reducción en la asignación del presupuesto. Por lo tanto, se requiere analizar la colección bibliográfica con base en indicadores cuantitativos para conocer las características y necesidades informativas de la comunidad. La evaluación cuantitativa es una herramienta que facilita al responsable bibliotecario la

\* Coordinación General de Bibliotecas, Universidad Autónoma de Yucatán, México  
pgomez@correo.uady.mx

\*\* Facultad de Ingeniería Química, Universidad Autónoma de Yucatán, México  
fernando.morales@correo.uady.mx  
paulina.martinez@correo.uady.mx  
rene.lopez@correo.uady.mx

toma de decisiones para la adquisición y descarte de material bibliográfico. Como propuesta, este trabajo utilizó un modelo de minería de datos para analizar los indicadores relacionados con: 1) número de volúmenes por subclase temática, 2) número de usos por subclase temática, y 3) número de volúmenes en uso por subclase temática. El modelo se validó con un estudio estadístico de las transacciones de préstamos tanto a domicilio como en sala registrados en una de las bibliotecas de la Universidad Autónoma de Yucatán durante el periodo de mayo de 2011 a junio de 2017.

**Palabras clave:** Evaluación de Colecciones; Indicadores Cuantitativos; Minería de Datos; Análisis Estadístico

### **Assesing university libraries through data mining**

*Carlos Antonio Pacheco Gómez, Luis Fernando Morales Mendoza, Paulina Martínez Isidro and Rene Lopez Flores*

#### **ABSTRACT**

Library collection development is a process that allows having the appropriate bibliographic material to meeting the needs of their users' community. In this process, one of the main restrictions is the reduction of budget allocation. Therefore, it is necessary to analyze the bibliographic collection based on quantitative indicators in order to know the characteristics and informative needs of the community. A quantitative assessment is a tool that makes easier for the librarian to take decisions for the acquisition or disposal of bibliographic material. As a proposal, this work used a data-mining model to analyze the indicators related to: (1) number of volumes by thematic subclass, (2) number of uses by thematic subclass, and (3) number of volumes in use by thematic subclass. A statistical study validated the model with loan transactions, both at home and in the reading room registered in one of the libraries of the Autonomous University of Yucatan during the period that went from May 2011 to June 2017.

**Keywords:** Library Assessment; Quantitative Indicators; Datamining; Statistical Analysis

## INTRODUCCIÓN

Las colecciones bibliográficas de una universidad, junto con el personal, constituyen uno de los elementos básicos que hacen posible los servicios bibliotecarios. Además, representan una importante proporción de los recursos financieros destinados a los servicios de información al usuario.

El presente trabajo aborda los retos que enfrentan las bibliotecas universitarias para el desarrollo de colecciones, particularmente la Universidad Autónoma de Yucatán (UADY) como un caso de estudio, y sus problemáticas asociadas con relación a la selección de materiales documentales para su adquisición o descarte. Se señala la importancia de la evaluación de sus colecciones para tomar mejores decisiones en el procedimiento de su desarrollo.

La correcta gestión de las colecciones tiene dos ventajas principales: 1) la rapidez en la obtención de las obras solicitadas, lo cual influye en la satisfacción de las necesidades informativas de los usuarios; 2) el ahorro económico, posibilitando la adquisición de mayor número de materiales y por consiguiente una mayor cobertura en cuanto a número de usuarios (Verela, 2000).

Para favorecer la evaluación de colecciones, se propone adaptar el modelo de minería de datos denominado Knowledge Discovery in Databases (KDD). Con su implementación se pretenden encontrar comportamientos predictivos o agrupaciones de elementos con características en común. Para conocer los detalles, el artículo aborda el tema de desarrollo de colecciones en bibliotecas universitarias y los retos asociados. Posteriormente, es descrito el modelo de minería de datos adaptado para la evaluación de colecciones. Los resultados son analizados y, finalmente, se elaboran algunas conclusiones y recomendaciones.

DESARROLLO Y EVALUACIÓN DE COLECCIONES EN BIBLIOTECAS  
UNIVERSITARIAS***Bibliotecas universitarias***

Las bibliotecas universitarias existen para “apoyar de una manera sistemática y efectiva las funciones básicas de la universidad a la que pertenecen, es decir, los planes y programas de docencia, investigación y difusión de la cultura universitarios” (Martínez Arellano, 1989). Por lo tanto, la forma como la biblioteca universitaria se hace indispensable es mediante la preservación y transmisión del conocimiento, reforzando el aprendizaje adquirido en las aulas (Lugo Hubp, 2000).

En cuanto a grandes cantidades de volúmenes que existen en las bibliotecas tenemos por ejemplo la Biblioteca del Congreso en EUA, con 34.5 millones, mientras que la Universidad de Harvard posee 16.8 millones de volúmenes, la Universidad de Yale 12.7 millones y la Universidad de Chicago 9.8 millones (ALA, 2017).

La necesidad por desarrollar, mantener y aumentar la calidad, cantidad y diversidad de información en los acervos ha sido una tarea importante en las bibliotecas universitarias de México. Como consecuencia de los avances científicos y tecnológicos, se enfrentan nuevos retos y oportunidades en el proceso de desarrollo de colecciones, que implican una concepción más amplia del proceso mismo y de las posibilidades asociadas a la formación y orientación de las colecciones (Negrete Gutiérrez, 1999).

En el caso de la UADY las bibliotecas se administran de manera centralizada, la institución cuenta en la actualidad con 12 bibliotecas: cinco de campus, dos multidisciplinarias, tres del nivel medio superior y dos departamentales. Su coordinador general, Alberto Arellano Rodríguez, recientemente declaró que no se tiene asignado un porcentaje fijo del total del presupuesto universitario. Los recursos se asignan dando prioridad a las bases de datos electrónicas y suscripciones de revistas, el recurso restante se destina para la compra de libros impresos (A. Arellano Rodríguez, comunicación personal, 13 de febrero de 2019). De esta manera, en 2014 se hicieron tres compras anuales, con un total de 8 000 volúmenes aproximadamente; en 2015 se hicieron dos compras anuales con un total de 4 447 volúmenes; en 2016 únicamente se realizó una compra de 1 424 volúmenes y en 2017 la adquisición fue de 1 102 volúmenes (Sistema Bibliotecario UADY, datos de 2017. <http://www.bibliotecas.uady.mx/>). Es evidente la disminución de recursos destinados para la adquisición de materiales documentales. Como consecuencia, aumenta la necesidad de tener información para tomar decisiones más acertadas en relación con la selección de material que se planea adquirir.

Por otro lado, parte importante en una biblioteca es la toma de decisiones referentes al descarte. Dado que la biblioteca es un organismo dinámico que está en continuo crecimiento, hay partes de la colección que van dejando de usarse en la misma medida en que los cambios y progresos científicos se van reflejando en la docencia e investigación realizadas por la institución (Verela, 2000). Por tanto, se puede decir que el descarte es consecuencia de las limitaciones de espacios en las bibliotecas y de la constante evolución del conocimiento humano (Mansilla y Verde, 2015).

Las dificultades para la adquisición y descarte de material motivan a los profesionales de la información para desarrollar herramientas de tipo intelectual con el objetivo de lograr una mayor eficacia en la construcción de colecciones. Una de estas herramientas es el proceso de desarrollo de colecciones.

### ***Desarrollo de colecciones***

Para Mosher (1982, citado en Mendoza Banda, 2007: 72), “desarrollo de colecciones es hacer una selección eficaz y oportuna de los materiales documentales para la formación cuidadosa de colecciones temáticas”. Evans (1995), uno de los autores más reconocidos e influyentes en el área, definió desarrollo de colecciones como el “proceso mediante el cual la biblioteca conoce y satisface, con certeza y rapidez, las necesidades de información de sus usuarios, utilizando los recursos de información generados dentro y fuera de la organización” (17).

De acuerdo con Evans (1995), el desarrollo de colecciones es un proceso continuo que se compone de seis etapas: 1) análisis de la comunidad, 2) políticas de desarrollo, 3) selección, 4) adquisición, 5) descarte y 6) evaluación. Un aspecto importante en el desarrollo de colecciones es fomentar su construcción de tal manera que satisfagan las necesidades de información de la comunidad. En este contexto, la evaluación proporciona las herramientas para lograr dicho objetivo.

La American Library Association (2017) define la evaluación de la colección como el “conjunto de estudios y operaciones que la biblioteca lleva a cabo para comprobar hasta qué punto la colección que ofrece responde a las necesidades de sus principales grupos de usuarios”. Esta definición destaca, más que el valor de las colecciones en sí mismas, su funcionalidad, la cual se entiende como el mayor o menor grado de adecuación a las necesidades de unos usuarios cuyos requerimientos de información se tienen previamente estudiados (Morales Fernández *et al.*, 2011).

Gómez Hernández (2002) comenta que evaluar una colección es “valorar la utilidad y pertinencia de las colecciones de una biblioteca con relación a sus usuarios o programas. [...] Es necesaria para enfocar los gastos de adquisiciones en los materiales más necesarios; conocer la colección de manera más efectiva”, además de que sirve de base para elaborar las políticas de desarrollo.

Para mantener colecciones idóneas es necesario realizar estudios, evaluaciones periódicas o ambas que avalen su calidad y su pertinencia en la biblioteca. El estudio de una colección implica su valoración en términos cuantitativos, cualitativos o ambos. Sus resultados son muy útiles en el proceso de selección o descarte de los materiales documentales. Con el estudio de la colección, el profesional encargado obtiene mediciones de indicadores y se familiariza con las obras que integran la colección y su grado de utilización (Morales Fernández *et al.*, 2011).

### ***Métodos de evaluación de colecciones***

Los métodos que se han ido desarrollando para la evaluación de colecciones incluyen los basados en la colección y en el uso, y a su vez, en indicadores cualitativos y cuantitativos.

*Métodos cuantitativos basados en la colección.* Para la aplicación de estos métodos se requiere tener una concepción clara de lo que se quiere medir, la manera de medirlo y la forma en la que se van a interpretar los resultados. Estos métodos bien aplicados tendrán como resultado información fiable y válida, originando con ello toma de decisiones más objetivas en temas tanto triviales como profundos de una biblioteca (Aguado, 2011).

*Métodos cuantitativos centrados en el usuario.* Las estadísticas de préstamos interbibliotecarios pudieran indicar deficiencias en áreas de la biblioteca, ya que son demandas de usuarios insatisfechas. Las estadísticas de uso indican si los materiales adquiridos están acordes a las necesidades de información de los usuarios, así como también determinar el monto de presupuesto y los descartes (Aguado, 2011).

*Métodos cualitativos basados en la colección.* Se apoyan en el supuesto de que los elementos de comparación con las colecciones propias responden a criterios objetivos que garantizan su valor como punto de referencia (Aguado, 2011). Para aplicar este método se seleccionan las listas de obras consideradas adecuadas de acuerdo con las características de la biblioteca y los usuarios. Posteriormente, se confronta con el catálogo para determinar en qué tanto por ciento las obras presentes en dichas listas se encuentran también en la colección (Massísimo i Sánchez, 2002).

La revisión *in situ* del contenido de los estantes con objeto de estudiar directamente dimensiones, alcance, profundidad y relevancia de la colección es un buen método, pero sólo aplicable a colecciones muy pequeñas y muy especializadas, para las cuales dispongamos de expertos temáticos.

*Métodos cualitativos centrados en usuarios.* Se basan en encuestas, en la formulación de preguntas mediante cuestionarios o entrevistas a los usuarios, elegidos al azar o seleccionados por su preferencia a determinado grupo considerado como representativo.

La *Tabla 1* muestra un resumen de los métodos cuantitativos y cualitativos basados tanto en la colección como en el uso.

Métodos	Basados en la colección	Basados en el uso
Cuantitativos	Tamaño/incremento de la colección Monto/aumento del presupuesto para materiales Estándares y fórmulas de tamaño de la colección	Estadísticas del préstamo interbibliotecario Estadísticas de circulación y uso <i>in situ</i> Estadísticas de disponibilidad en el estante
Cualitativos	Listas de control Estudios de verificación Observación de usuarios Control directo de la colección Descripción de la colección, con asignación de niveles	Encuestas de opinión de usuarios Análisis de citación Grupos focales

Tabla 1. Métodos de análisis de colección

Fuente: Aguado (2011)

### ***Análisis de datos***

Tradicionalmente las bibliotecas han almacenado información acerca de las colecciones y los servicios que prestan; sin embargo, con algunas excepciones, dicha información no es sometida a procesos que permitan obtener información de apoyo en la toma de decisiones para la gestión bibliotecaria (Blake y Schleper, 2004).

Con los sistemas automatizados de las bibliotecas, esta situación cambió y, en la actualidad, una cantidad importante de bibliotecas universitarias cuenta con una base de datos en sus sistemas factible de ser utilizada como fuente de información. Sin embargo, estos beneficios de la automatización parecen haber sido utilizados sólo para la asistencia informativa en la toma de decisiones de nivel operativo, desaprovechándolo para la toma de decisiones de nivel directivo (Aguado *et al.*, 2007).

Por otro lado, tenemos que el crecimiento de las bases de datos y el aumento de las capacidades de almacenamiento de información han hecho que puedan tener a disposición un gran banco de datos de sus actividades diarias. En muchas de estas organizaciones han comenzado a utilizarlas para el apoyo a la gestión. Su estudio permite tener un panorama retrospectivo y con ello determinar una tendencia (Candás Romero, 2006).

El presente trabajo considera que la integración de la bibliotecología con sistemas de información y análisis de datos favorece dos aspectos del

desarrollo de colecciones: por un lado, conocer áreas poco desarrolladas y por consiguiente que necesitan ser priorizadas; por otro, reconocer áreas que no tengan utilidad en la colección y ocupan espacio en las estanterías.

### MODELO DE MINERÍA DE DATOS PARA LA EVALUACIÓN DE COLECCIONES

El diseño de la metodología es de carácter estadístico y de estudio de caso. El caso de estudio es evaluado con los indicadores seleccionados. El modelo de evaluación identificado utilizó como base el procedimiento metodológico de minería de datos KDD (Knowledge Discovery in Databases), que incluye las siguientes fases: Parámetros de entrada, Selección de datos, Procesamiento, Transformación, Análisis de datos y Presentación de resultados (Fayyad, Piatetsky-Shapiro y Smyth, 1996). La *Figura 1* presenta cada una de las fases del modelo de minería de datos aplicado al desarrollo de colecciones. Posteriormente, se describen las actividades en cada etapa.

#### *Modelo de minería de datos*



*Figura 1.* Modelo de evaluación de colecciones identificado  
Fuente: elaboración propia

#### Fase 1. Establecimiento de parámetros de entrada

- Indicar los criterios que se desean evaluar, que cubran lo establecido en la “Política general de desarrollo de colecciones”
- Determinar el área de conocimiento que se desea evaluar.
- Identificar qué sistema gestiona la biblioteca, credenciales de acceso tanto del sistema como de la base de datos.
- Generales: año o periodo a evaluar.

#### Fase 2. Selección de datos

- Determinar la fuente de datos principal físicamente, es decir, la base de datos que gestiona el acervo bibliográfico, así como las transacciones que realiza la biblioteca.



- Identificación y uso de las herramientas de administración de bases de datos, dicha herramienta deberá permitir tener una conexión directa a la base de datos y tener la capacidad de realizar consultas estructuradas.
- Selección y extracción de las tablas implicadas para identificar las tablas que almacenan la información referente a:
  - Información de la biblioteca,
  - Acervo bibliográfico, y
  - Registros de los préstamos a domicilio y en sala.

#### Fase 3. Procesamiento

- Diseño de consultas SQL que arrojen vistas de conjuntos de los criterios evaluados.
- Seleccionar todos los volúmenes de libros que se encuentren en la biblioteca elegida para el estudio.
- Seleccionar todos los préstamos de libros en la biblioteca de estudio durante el rango de fecha.
- Conjuntos de: 1) Libros clasificados, 2) Préstamos a domicilio, 3) Préstamos a sala, 4) Registro en sala, y 5) Acervos bibliográficos.

#### Fase 4. Transformación

- Adecuar la información de los campos de las tablas de tal forma que permita realizar los conteos y cálculos de los indicadores cuantitativos de uso de la colección.
- Normalizar campos implicados, particularmente los que se refieran a la clasificación de los libros.
- Seccionar la clasificación de los volúmenes de tal forma que se puedan crear subclases de áreas de conocimiento, utilizando la Clasificación de la Biblioteca del Congreso (LC).
- Diseñar y codificar algoritmos que permitan hacer los cálculos complejos, las cuales no se puedan realizar con ayuda del administrador de la base de datos o con una hoja de cálculo.

#### Fase 5. Análisis de datos

- Analizar sobre los listados generados en la fase anterior los siguientes indicadores: 1) Materiales en la colección por usuario, 2) Materiales prestados en sala *per cápita*, 3) Materiales prestados *per cápita*, 4) Índice de circulación, 5) Porcentaje de adquisiciones por clase temática, 6) Número de materiales por clase temática, 7) Índice de uso por clase temática, 8) Porcentaje de uso por clase temática, 9) Porcentaje de la colección no utilizada, 10) Tasa de crecimiento de la colección.

#### Fase 6 Presentación de resultados

Presentación de los resultados a las partes interesadas, particularmente al

responsable de desarrollo de colecciones. Su sugiere utilizar herramientas gráficas que permitan tener una visión clara de los resultados obtenidos en el estudio.

### ***Caso de estudio***

El modelo seleccionado se validó utilizando los registros de la biblioteca del Campus de Ciencias Exactas e Ingenierías (CCEI) de la UADY. El número total de volúmenes analizados fue 38 003. Para el uso de la colección se utilizaron las transacciones de volúmenes usados, lo que corresponde a 1 229 233 operaciones de préstamos a domicilio y préstamos en sala. El periodo de análisis se limitó a las operaciones de mayo de 2011 a junio de 2017.

Los indicadores cuantitativos definidos como los criterios evaluados están clasificados por el número de volúmenes y su uso. Los criterios son: 1) Número de materiales por clase temática, 2) Porcentaje de uso por clase temática, 3) Porcentaje de volúmenes utilizados por clase temática, y 4) Porcentaje de la colección no utilizada. Para la selección de los datos se utilizaron las tablas de acervo bibliográfico, préstamos a domicilio y préstamos en sala.

Para los cálculos cuantitativos por área temática se optó por crear una tabla de subclases temáticas con base en los datos que permitiera realizar los cálculos de su uso. La tabla se alimentó con base en las clases y subclases del sistema de clasificación desarrolladas por la Biblioteca del Congreso de Estados Unidos de América (<http://www.loc.gov/catdir/cpsol/lcco/>).

*Indicador número de volúmenes por subclase temática.* Es el número de volúmenes de libros agrupados por subclase temática pertenecientes a la biblioteca elegida. El diagrama de flujo de la *Figura 2* describe el proceso para el cálculo. Con el procedimiento se creó una variable vector en memoria, de tal forma que contenga la información de la subclase y una variable como contador para registrar el número de ocurrencias de la subclase en la colección de la biblioteca de CCEI.

Para calcular el número de usos (préstamos a domicilio y préstamos en sala) por subclase temática se empleó un procedimiento similar, la diferencia es la consulta a la base datos. En este caso, se construyó una consulta que permitiera listar el identificador y la clasificación de todos los libros que fueron prestados en la biblioteca del CCEI contenidos en el periodo establecido. *Indicador número de volúmenes en uso por subclase temática.* Se utilizó el vector con el indicador número de volúmenes por subclase temática, y se itera en

cada uno de sus elementos. Por cada uno de ellos se realizó una consulta a la base de datos para recuperar los elementos que forman parte de la subclase y que además hayan sido usados al menos una vez en el periodo de estudio.

En la consulta anterior, si al menos se encontrara un préstamo del material, se contabiliza en un vector para que finalmente se contabilicen los volúmenes que fueron utilizados. La *Figura 3* representa la operación del procedimiento mediante un diagrama de flujo.

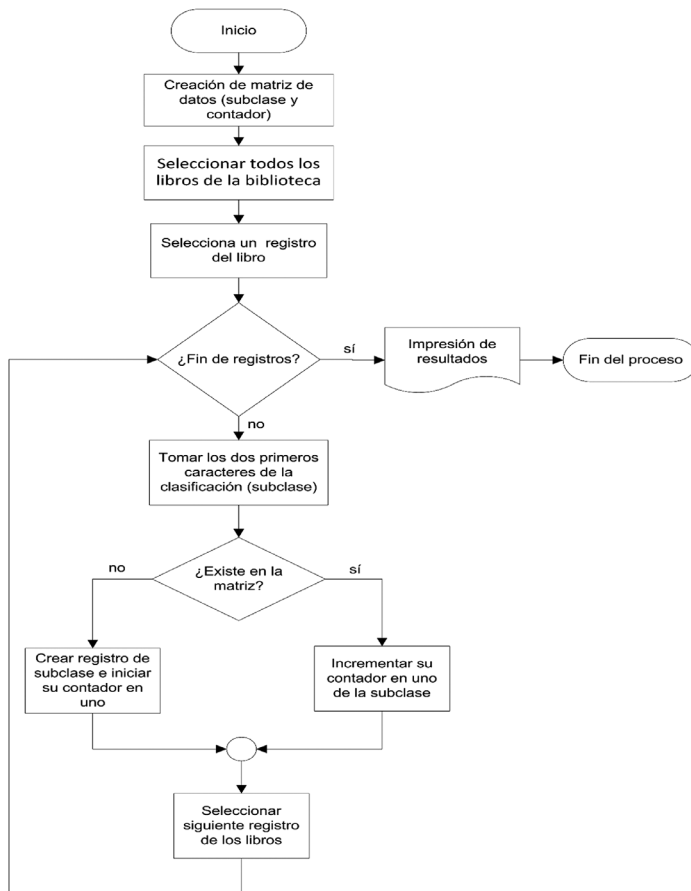


Figura 2. Diagrama de flujo para número de volúmenes por subclase temática

Fuente: elaboración propia

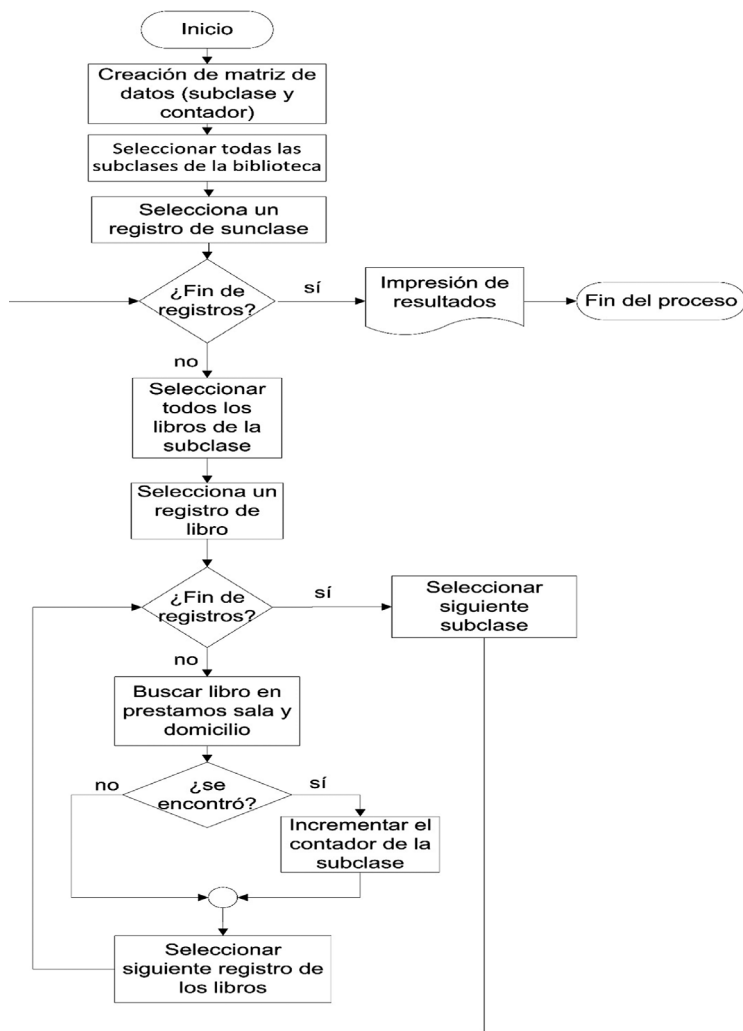


Figura 3. Diagrama de flujo para número de volúmenes en uso por subclase temática  
Fuente: elaboración propia

Un ejemplo de la estructura de datos final que se construyó en memoria con el número de volúmenes en uso por subclase temática se muestra en la *Tabla 2*. En el *Anexo* se localiza la descripción para las subclases presentadas en las tablas de resultados.

Subclase	Volúmenes	Uso colección	Volúmenes utilizados
QA	12144	147615	6961
QD	1925	40230	1320
TA	3927	39557	1793
QC	1805	31280	1209
TJ	1124	19396	677
TP	1420	16308	746

Tabla 2. Ejemplo de resultados calculados para los volúmenes utilizados

Con el cálculo de los tres indicadores, el siguiente paso es analizar los datos. *Análisis de los datos.* Para esta fase se empleó una hoja de cálculo de Excel en la cual se utilizaron los vectores recuperados para cada indicador para realizar los cálculos relacionados con porcentaje por volúmenes, uso, volúmenes usados y factor de uso para cada subclase. Para el cálculo de los factores se emplean las funciones descritas a continuación.

El *porcentaje de volúmenes utilizados* se calculó mediante la fórmula

$$Pv = \frac{Vol.}{\sum_{i=1}^n Vol. [i]} \quad (1)$$

donde:

$Pv$  = porcentaje de volúmenes utilizados

$Vol.$  = número de volúmenes de la subclase

El *porcentaje de uso sobre el total* se calculó mediante la fórmula

$$Pu = \frac{Uso}{\sum_{i=1}^n Usos[i]} \quad (2)$$

donde:

$Pu$  = porcentaje de uso de la subclase

Finalmente, el factor de uso ( $fu$ ) se define con la fórmula

$$f_u = \frac{\text{Uso de la subclase}}{\text{Volúmenes de la subclase}} \quad (3)$$

### RESULTADOS Y DISCUSIÓN

En este apartado se exponen los resultados encontrados en el caso de estudio. Se presentarán gráficos Pareto para delimitar las áreas vitales de la colección, gráficos de comparación entre número de volúmenes y sus usos y las estadísticas que se realizaron sobre los datos encontrados.

En junio de 2017 el Sistema Bibliotecario de la UADY contabilizó 701 967 piezas documentales en todas sus colecciones, de las cuales 381 629 volúmenes corresponden a libros, lo cual representa 54 %. Para el caso de la biblioteca del CCEI, se encontró que para la colección de libros fueron en total 38 003 volúmenes, es decir, 10 % de la colección total del Sistema Bibliotecario. En cuanto a la distribución temática de la biblioteca del CCEI se realizó un análisis de Pareto, dando como resultado 80 % los datos presentados en la *Tabla 3*.

Subclase	Volúmenes	% colección	% acumulado
QA	12144	32 %	32 %
TA	3927	10 %	42 %
QD	1925	5 %	47 %
QC	1805	5 %	52 %
TK	1703	4 %	57 %
HD	1449	4 %	60 %
TP	1420	4 %	64 %
TJ	1124	3 %	67 %
TD	1034	3 %	70 %
Q	1012	3 %	72 %
T	1010	3 %	75 %
HF	859	2 %	77 %
TH	677	2 %	79 %
TS	655	2 %	81 %

*Tabla 3.* Pareto distribución de la colección en CCEI

Para facilitar la lectura de la información se presenta la *Figura 4*. Este análisis de Pareto permite identificar las subclases más relevantes que conforman la

colección de la biblioteca. Se encontró que 13 subclases temáticas de un total de 136 representan 80 % del total de la colección de libros de la biblioteca del CCEI, es decir, 9,5 % del total de subclases. Como se puede observar, la subclase QA (Matemáticas) es la que cuenta con mayor desarrollo, con 32 % de toda la colección; esto evidentemente refleja las características de una biblioteca de ciencias exactas e ingenierías, al igual que el resto de las subclases, todas pertinentes a las áreas temáticas de los planes de estudio de las comunidades de usuarios.

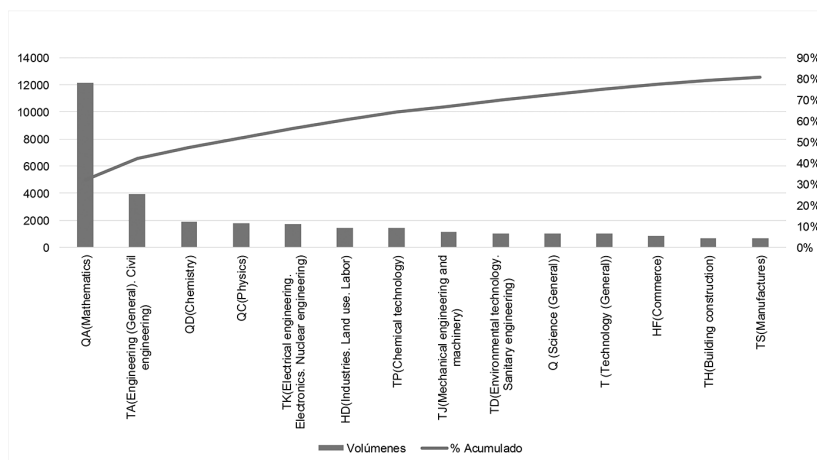


Figura 4. Análisis Pareto de volúmenes por subclase temática

Por otro lado, en cuanto a la distribución de los usos por clases temáticas el análisis de Pareto da como resultados los datos presentados en la *Tabla 4*.

Subclase	Uso colección	% de uso	% Acumulado
QA	483263	39.3 %	39.3 %
QD	136265	11.1 %	50.4 %
TA	132206	10.8 %	61.2 %
QC	103442	8.4 %	69.6 %
TJ	66496	5.4 %	75.0 %
TP	56107	4.6 %	79.5 %
TK	52072	4.2 %	83.8 %

Tabla 4. Pareto uso de la colección

La *Figura 5* permite identificar las subclases más utilizadas por la comunidad. Con base en el análisis de Pareto, tomando como criterio el uso de la

colección, se deduce que las subclases temáticas con mayor relevancia para la biblioteca de CCEI son Matemáticas, Química, Ingeniería Civil, Física, Ingeniería Mecánica y Maquinaria, Tecnología Química, e Ingeniería Electrónica. Los temas corresponden con el tipo de planes de estudio que se atienden en el campus donde se encuentra la biblioteca.

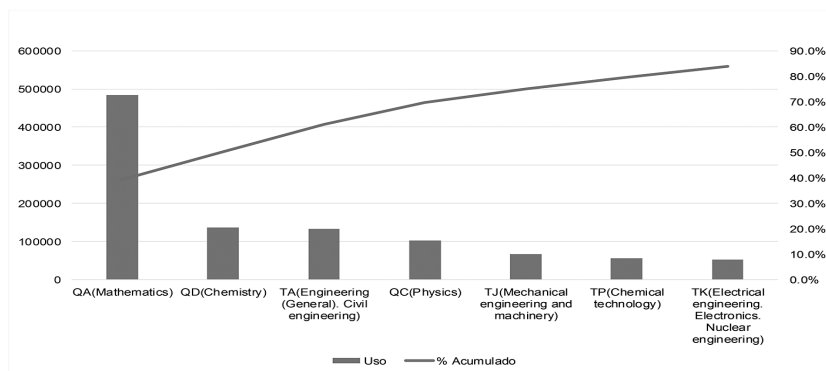


Figura 5. Análisis Pareto uso de la colección

Con base en el análisis de Pareto sobre el uso de la colección se determinó que siete subclases temáticas de un total de 136 representan 80 % de las subclases más utilizadas por la comunidad de la biblioteca del CCEI, es decir, 5.1% del total de subclases. Las subclases temáticas que se muestran con mayor porcentaje de uso sugieren ser prioridad en la actividad relacionada a desarrollo de colecciones. Son las subclases temáticas importantes para las carreras y diferentes actividades académicas del CCEI, por lo tanto, estas siete subclases tendrían mayor ponderación al seleccionar materiales para adquirir.

En relación con el análisis estadístico, los datos de la tabla generada en la fase de análisis de datos se ordenaron de mayor a menor respecto al factor de uso, resultado de dividir el número de usos entre el número de volúmenes, dando como resultado la *Tabla 5*, que muestra los primeros 11 resultados.

Subclase	Volúmenes	% volúmenes	Uso colección	% Uso colección	Vol. Usados	% de Vol. usados	Factor de uso
QA	12144	32.0 %	483263	39.3 %	9380	77 %	39.8
TA	3927	10.3 %	132206	10.8 %	2748	70 %	33.7
QD	1925	5.1 %	136265	11.1 %	1721	89 %	70.8
QC	1805	4.7 %	103442	8.4 %	1541	85 %	57.3
TK	1703	4.5 %	52072	4.2 %	1223	72 %	30.6
HD	1449	3.8 %	21455	1.7 %	1060	73 %	14.8



TP	1420	3.7 %	56107	4.6 %	1157	81 %	39.5
TJ	1124	3.0 %	66496	5.4 %	939	84 %	59.2
TD	1034	2.7 %	8323	0.7 %	727	70 %	8.0
Q	1012	2.7 %	12919	1.1 %	772	76 %	12.8
T	1010	2.7 %	24639	2.0 %	710	70 %	24.4

Tabla 5. Factor de uso por subclase

La distribución de los usos respecto de las subclases temáticas está relacionada al número de volúmenes, debido a que se observa mayor uso donde hay mayor volumen si bien con ciertas excepciones, las cuales evidencian algún punto crítico deseable de un estudio detallado por parte de un bibliotecario profesional, por ejemplo, correlacionar las subclases con poco uso y muchos volúmenes con la obsolescencia de sus títulos.

En la *Figura 6* se puede observar una gráfica comparativa entre los volúmenes de las subclases y sus respectivos usos.

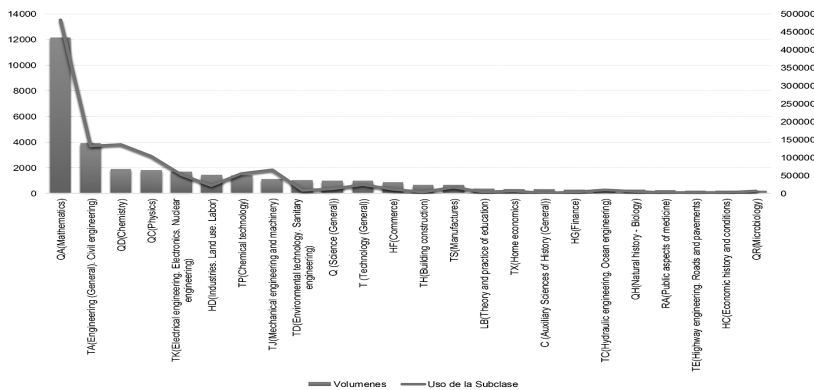


Figura 6. Análisis de volúmenes por uso

Al comparar proporcionalmente los volúmenes con los usos de las subclases se observaron los siguientes puntos críticos:

- QD (Química) Mucho uso pocos volúmenes
- QC (Física) Mucho uso pocos volúmenes
- HD (Industrias, Uso del Suelo) Poco uso y muchos volúmenes
- TJ (Ingeniería Mecánica) Mucho uso pocos volúmenes
- TD (Medio Ambiente) Poco uso y muchos volúmenes
- TH (Construcción de Edificios) Poco uso y muchos volúmenes

Los puntos críticos mencionados son indicios de una posible obsolescencia de la colección, que no se da suficiente difusión al material bibliográfico, que las solicitudes de los cuerpos académicos no corresponden a los planes de estudio, los títulos no corresponden a las necesidades de la comunidad o al uso de recursos electrónicos. De lo anterior se puede establecer que los puntos críticos observados en la evaluación son resultado de un problema multifactorial que amerita un análisis posterior.

En relación con el descarte de material, se analizó la información de la columna % de vol. utilizados por subclase de la *Tabla 5* para determinar lo siguiente:

- QA (Matemáticas) con 12 144 volúmenes tiene un porcentaje de no utilización de 23 %, es decir, 2 764 volúmenes.
- TA (Ingeniería General. Ingeniería Civil) con 3 927 volúmenes tiene un porcentaje de no utilización de 30%, es decir, 1 179 libros.

Dada la cantidad de volúmenes no utilizados de estas clases es conveniente realizar una exploración más exhaustiva para determinar las causas. Se propone volver a ejecutar el modelo con estos nuevos parámetros de entrada para que se analicen a un nivel más elevado en granularidad de las subclases temáticas para con ello acotar las secciones no utilizadas y tener mayor certeza en el descarte de materiales.

## CONCLUSIONES

Se validó el modelo seleccionado para la evaluación de una colección bibliográfica con el apoyo de herramientas estadísticas y uso de gráficos, utilizando información de la biblioteca del CCEI de la UADY. Siguiendo las fases del modelo se realizó una evaluación cuantitativa sobre el uso de la colección. La aplicación de la metodología proporcionó resultados que pueden ser útiles para la toma de decisiones en la selección y en el descarte de títulos.

A partir del análisis de número de volúmenes por su uso se observó, de manera general, que la colección bibliográfica de la biblioteca del CCEI está distribuida de tal forma que cubre con la demanda de su comunidad. Sin embargo, el análisis también reveló puntos críticos de la colección, es decir, subclases con mucha demanda y poco desarrollo (como Química o Física). Por el contrario, también existen subclases con poco uso y muchos volúmenes (por ejemplo, Medio Ambiente o Construcción de Edificios).

Para el descarte de materiales se observaron porcentajes no utilizados de cada subclase. Se propone como trabajo futuro, para aquellas clases cuyo porcentaje de no utilización sea elevado, realizar un análisis más profundo de la subclase. Una propuesta es añadir más parámetros de entrada al modelo, específicamente granular a un nivel inferior la clasificación de los libros y ejecutar de nuevo el modelo, lo que permitiría acotar específicamente que áreas de la subclase no son utilizadas y tener un análisis más detallado de la colección.

Finalmente, se recomienda completar el análisis cuantitativo desarrollado en este trabajo con un estudio cualitativo sobre el comportamiento de la comunidad de usuarios. Con ambos resultados es posible obtener elementos más confiables que ayuden en la toma de decisiones para el desarrollo de las colecciones y con ello utilizar eficientemente el presupuesto.

#### *Agradecimientos*

Al Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP) por el apoyo con el proyecto Folio UADY-PTC-228.

#### REFERENCIAS

- Aguado, A., M. Sandra, C. Archuby, M. Fushimi, C. González, M. Pichinini, C. Rozemblum, y M. Pené. 2007. "Desarrollo de colecciones en bibliotecas universitarias Metodología de evaluación". Proyecto de investigación acreditado y financiado por Universidad Nacional de La Plata, Argentina (H432) Período 2006-2007. Fecha de consulta: 4 de noviembre de 2018. [http://eprints.rclis.org/9520/1/Aguado%2C\\_A.\\_y\\_otros\\_Evaluacion\\_de\\_colecciones\\_en\\_bibliotecas\\_universitarias.pdf](http://eprints.rclis.org/9520/1/Aguado%2C_A._y_otros_Evaluacion_de_colecciones_en_bibliotecas_universitarias.pdf)
- Aguado, A. 2011. *Gestión de Colecciones*. Argentina: Alfagrama Ediciones.
- ALA (American Library Association). 2017. *The Nation's Largest Libraries: A Listing By Volumes Held*. Fecha de consulta: 7 de mayo de 2017. <http://www.ala.org/tools/libfactsheets/alalibraryfactsheet22>
- Blake, J. C. y S. P. Schleper. 2004. "From data to decisions. Using surveys and statistics to make collection management decisions". *Library Collections, Acquisitions, & Technical Services*, 460-464.
- Candás Romero, J. 2006. "Minería de datos en bibliotecas: bibliominería". *BiD: textos universitaris de biblioteconomia i documentació* (17). Fecha de consulta: 4 de noviembre de 2016. <http://bid.ub.edu/17canda2.htm>
- Evans, E. 1995. *Developing library collections*. 3th. Littleton, Colorado: Libraries Unlimited.
- Fayyad, U., G. Piatetsky-Shapiro y P. Smyth. 1996. "From Data Mining to Knowledge Discovery in Databases". *American Association for Artificial Intelligence*, 37-54.
- Fuentes, J. 1999. *Evaluación de bibliotecas y centros de documentación e información*. Gijón, Asturias: Ediciones Trea.

- Gómez Hernández, J. 2002. *Gestión bibliotecaria*. Fecha de consulta: 20 de marzo de 2017. <https://es.scribd.com/doc/58586517/Gestion-de-Bibliotecas-Gomez-Hernandez-2002>
- Lugo Hubp, M. 2000. "Las bibliotecas universitarias mexicanas: apuntes para un diagnóstico". *Métodos de Información* 7 (40): 45-53.
- Mansilla, G. y M. Verde. 2015. "Descarte de documentos: una propuesta metodológica para bibliotecas". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 29 (67): 91-111. Fecha de consulta: 10 de marzo de 2017. <http://revistas.unam.mx/index.php/ibi/article/download/54477/48452>
- Martínez Arellano, F. 1989. *Políticas de selección y adquisición de material bibliográfico en las instituciones de educación superior*. México: UNAM, ANUIES.
- Massísimo i Sánchez, A. 2002. "Evaluación de colecciones en las bibliotecas universitarias (i). Métodos basados en el estudio de la colección." *Anales de documentación* (5): 245-272.
- Mendoza Banda, R. 2007. "El desarrollo de colecciones en bibliotecas universitarias de México: panorama actual, problemática y posibles soluciones". Tesis de licenciatura, Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras. <https://repositorio.unam.mx/contenidos/440392>
- Morales Fernández, T., O. Águila García, A. Diago Gómez y L. Clavero Fleites. 2011. "Guía metodológica para la evaluación de colecciones: experiencia con diez especialidades médicas". Trabajo presentado en Satellite Conference: Acquisition and Collection Development Section, Las Islas Vírgenes de los Estados Unidos, 10-11 de agosto.
- Negrete Gutiérrez, M. D. 1999. "El impacto del cambio en el desarrollo de colecciones en bibliotecas universitarias". *Scire: representación y organización del conocimiento* 5 (1): 55-64.
- Verela, Xose. 2000. "Programa de gestión y desarrollo de colecciones de una biblioteca universitaria (II). Política de gestión de colecciones." *Boletín de la Asociación Andaluza de Bibliotecarios* 58: 14. <http://eprints.rclis.org/5978/1/baaba58a3.pdf>

## ANEXO

### Lista de subclases de la clasificación LC

TH	Building construction
TP	Chemical technology
QD	Chemistry
HF	Commerce
TK	Electrical engineering. Electronics. Nuclear engineering
TA	Engineering (General). Civil engineering
TD	Environmental technology. Sanitary engineering
HD	Industries. Land use. Labor

TS	Manufactures
QA	Mathematics
TJ	Mechanical engineering and machinery
QC	Physics
Q	Science (General)
T	Technology (General)

*Para citar este texto:*

Pacheco Gómez, Carlos Antonio, Luis Fernando Morales Mendoza, Paulina Martínez Isidro y Rene Lopez Flores. 2019. "Evaluación de colecciones en una biblioteca universitaria utilizando minería de datos". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 33 (81): 201-221.

<http://dx.doi.org/10.22201/iibi.24488321xe.2019.81.58058>