

1. Título del proyecto

Modelo de datos orientado a grafos para la recomendación de libros de una biblioteca pública

2. Área de Investigación

Área de investigación	Línea de Investigación	Disciplina OCDE
sistemas	base de datos	

3. Duración del proyecto (meses)

12

4. Tipo de proyecto

Individual	<input type="radio"/>
Multidisciplinario	<input type="radio"/>
Director de tesis pregrado	<input checked="" type="radio"/>

4. Datos de los integrantes del proyecto

Apellidos y Nombres	coyla idme elmer
Escuela Profesional	Ingeniería de Sistemas
Celular	979950909
Correo Electrónico	ecoyla@unap.edu.pe

- I. **Título** (El proyecto de tesis debe llevar un título que exprese en forma sintética su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25)

Modelo de datos orientado a grafos para la recomendación de libros de una biblioteca pública

- II. **Resumen del Proyecto de Tesis** (Debe ser suficientemente informativo, presentando -igual que un trabajo científico- una descripción de los principales puntos que se abordarán, objetivos, metodología y resultados que se esperan)

La investigación abordará el tema que involucra las propiedades únicas de los modelos de base de datos orientada a grafos para el almacenamiento y procesamiento de información compleja aplicadas a la recomendación de información bibliográfica, debido que las bibliotecas públicas usan formatos estándares para catalogar su información y gestionándolas mediante sistemas de información bibliográfica diseñados precisamente a partir de estos formatos, sin embargo, estos softwares no permiten proporcionar sugerencias de libros a los usuarios, a causa de las limitaciones en cuanto al almacenamiento y procesamiento de información altamente conectada requeridas por procedimientos propios de la generación de recomendaciones y que padecen los modelos de base de datos tradicionales sobre los cuales marchan estos sistemas.

El objetivo principal de la investigación es usar un modelo de base de datos orientado a grafos para proporcionar recomendaciones de libros que obtenga variedad en los resultados, aplicando el filtrado basado en contenido como técnica de recomendación, eventualmente considerando los formatos estándar manejados por las bibliotecas públicas. Se plantea una investigación con enfoque cuantitativo de alcance experimental, así mismo, se espera obtener resultados competitivos expresados en la variedad de las recomendaciones. Para desarrollar el proyecto se utilizarán los datos bibliográficos de la Biblioteca Municipal de Puno.

III. Palabras claves (Keywords) (Colocadas en orden de importancia. Máx. palabras: cinco)

Base de datos orientado a grafos, recomendación de libros, filtrado basado en contenido, biblioteca pública

IV. Justificación del proyecto (Describa el problema y su relevancia como objeto de investigación. Es importante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar)

Las bibliotecas con el fin de catalogar e intercambiar información han codificado sus materiales bibliográficos y audiovisuales con ayuda de formatos específicos estandarizados, "el formato MARC es más utilizado en el mundo y un estándar de facto, la mayoría del software de gestión de bibliotecas disponibles utiliza variaciones de este formato" (International Federation of Library Associations and Institutions IFLA, 2017).

En Latinoamérica las bibliotecas han optado por utilizar software desarrollado por terceros con la finalidad de adaptar sus procesos a estándares internacionales, en el estudio de Alarcón Escriba, (2020) mostró que los sistemas de gestión de bibliotecas basadas en software libre son usadas en gran mayoría por bibliotecas universitarias y públicas. Estos sistemas de información marchan sobre la variante del formato MARC.

Las bibliotecas públicas locales resguardan gran cantidad de información valiosa usando software de gestión bibliográfica de acceso libre basados en el formato de registro UNIMARC, evidentemente este modelo ha mostrado un desempeño importante en la automatización del acceso al público de la información bibliográfica. Sin embargo, esta información organizada además de ser útil para facilitar el acceso, también puede ser explotada con el propósito de proporcionar recomendaciones.

Los sistemas de gestión bibliotecaria utilizadas por bibliotecas públicas no permiten proporcionar recomendaciones de libros a sus usuarios, esta situación empeora si se trata de sugerir libros personalizados a tiempo real. Sin embargo, la recomendación de libros fue tratada utilizando otros métodos y modelos de datos que difieren de las tradicionales, así mismo, en todos los casos el formato estandarizado manejado por bibliotecas públicas no fue considerado en lo absoluto.

Una de las causas del problema es porque los sistemas bibliotecarios funcionan sobre modelos de bases de datos relacionales, los cuales presentan dificultades al procesar eficientemente consultas complejas y concurrentes de filtrado que involucran la alta relación de datos, propios de la generación de recomendaciones, así mismo, los gestores de bases de datos relacionales necesitan usar una gran cantidad de uniones para satisfacer las consultas que involucran la relación compleja de datos. Según Velásquez Vargas (2014) los rendimientos de las bases de datos relacionales se deterioran en las consultas intensivas y con procesamientos de datos muy altos. Así mismo, Medhi y Baruah (2017) y reafirmado por Rodríguez Reyes (2020) encontraron que las bases de datos relacionales son más lentas ejecutando consultas complejas frente a una basada en un modelo de datos no relacional.

Estas restricciones del modelo de datos hacen inviable la implementación de "herramientas y técnicas de software que brinden sugerencias de elementos que pueden ser de utilidad para un usuario" (Ricci et al., 2011), provocando que la información resguardada en forma de libros queden invisibles y desconocidos, con un futuro evidentemente de olvido permanente de sus conocimientos por parte de la sociedad.

En este trabajo de investigación se propone una alternativa de solución mediante el uso de un modelo de base de datos orientado a grafos como método para procesar recomendaciones de libros, usando el filtrado basado en contenido como principal técnica de recomendación, Mendoza Olgún et al. (2019) señalan que esta técnica implica recomendaciones basados en las similitudes de los ítems y pueden realizarse usando calificaciones que han recibido los objetos de forma anónima o bien considerando sus características. Así también de prestar especial importancia al formato de catalogación estándar como fuente de representación de los datos.

V. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para evidenciar el conocimiento existente y el aporte de la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo)

Amézquita Llerena (2020) utilizó un gestor de base de datos orientado a grafos para gestionar y procesar información de un sistema de recomendaciones relacionada a la medicina. Representó los nodos "enfermedades y síntomas" en una base de datos de grafos, conectándolos entre ellas a partir de síntomas, enfermedades y sus relaciones, información obtenida de dos repositorios médicos reconocidos. Se centró en generar la recomendación más precisa. Concluyó que un gestor de base de datos es ideal para modelos estructuradas con un alto nivel de relación de datos, logró resultados complejos con consultas simples

VI. Hipótesis del trabajo (Es el aporte proyectado de la investigación en la solución del problema)

La implementación del modelo de base de datos orientado a grafos recomienda libros de una biblioteca pública con resultados variados.

VII. Objetivo general

Implementar un modelo de base de datos orientado a grafos para recomendar libros de una biblioteca pública que obtenga resultados variados.

VIII. Objetivos específicos

Representar datos de libros de una biblioteca pública en una base de datos orientado a grafos.
Implementar algoritmos de similitud entre nodos de libros representados en la base de datos orientada a grafos.
Implementar consultas de filtrado basados en contenido en la base de datos orientada a grafos.

IX. Metodología de investigación (Describir el(los) método(s) científico(s) que se empleará(n) para alcanzar los objetivos específicos, en forma coherente a la hipótesis de la investigación. Sustentar, con base bibliográfica, la pertinencia del(los) método(s) en términos de la representatividad de la muestra y de los resultados que se esperan alcanzar. Incluir los análisis estadísticos a utilizar)

Para esta investigación se utilizará la técnica de recomendación basada en contenido con retroalimentación implícita, debido a la ausencia de perfiles de usuario. El conjunto de datos a utilizar para realizar las pruebas de recomendación será obtenido del catálogo en línea de la Biblioteca Municipal de Puno. Se seleccionarán aleatoriamente dos grupos de usuarios, los primeros recibirán las recomendaciones y los segundos no, en orden de determinar la variedad de los resultados de la recomendación se comparan el tráfico de libros vistos generados por estos grupos. En este sentido, de acuerdo al propósito y objetivos formulados, esta investigación reúne las condiciones suficientes para ser una investigación cuantitativa, así mismo, el diseño de investigación es de alcance experimental porque la hipótesis sugiere la realización de un experimento. Según Ríos Ramírez (2017) los diseños experimentales busca describir los efectos de una intervención, estímulo o causa de un hecho. Esto implica la manipulación de la variable independiente para observar sus efectos sobre la variable dependiente. De acuerdo a Hernández Sampieri (2014) si un grupo recibe el tratamiento experimental y el otro no, se le denomina diseño con posprueba únicamente y grupo de control. Los cuales encajan perfectamente con el tratamiento propuesto por esta investigación.

X. Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

Alarcón Escriba, M. J. (2020). Estado del arte de la automatización de bibliotecas universitarias mediante el uso de Sistemas Integrados de Gestión Bibliotecaria (SIGB) de código libre. Universidad Nacional Mayor de San Marcos.
Amézquita Llerena, J. F. (2020). Software de recomendación médico basado en modelo de datos orientado a grafos con Neo4j. Universidad Católica de Santa María.
Dharmawan, I. N. P. W., & Sarno, R. (2017). Book recommendation using Neo4j graph database in BibTeX book metadata. 3rd International Conference on Science in Information Technology:

XI. Uso de los resultados y contribuciones del proyecto (Señalar el posible uso de los resultados y la contribución de los mismos)

Los resultados pueden usarse como fuente para implementar sistemas de recomendaciones sin mayores complicaciones técnicas y aplicarlas a cualquier ámbito del conocimiento, en especial a la información administrada por las bibliotecas públicas, a fin de promocionar sus libros con información valiosa a la sociedad. Los sistemas de recomendación se caracterizan por la capacidad de manipular grandes volúmenes de datos y a menudo se prioriza solo algoritmos de filtrado o modelos de inteligencia artificial, dejando en segundo orden el modelo de datos sobre el cual existen los datos, sin embargo, este proyecto pretende dejar en evidencia que estas prioridades no necesariamente son en ese orden, revelando que sí se optimiza la incorporación de información en modelos de datos adecuados puede mejorar la eficiencia y reducir la complejidad del sistema.

XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

Innovación en el uso del modelo de datos orientado a grafos para desarrollar sistemas de recomendación, aplicadas sobre la información bibliográfica resguardada en bibliotecas públicas.

ii. Impactos económicos

Al generar recomendaciones mediante el uso de una base de datos de grafos, las bibliotecas públicas podrán usar este modelo para abaratar los costos que requeriría implementar un sistema de recomendación que promueva sus libros.

iii. Impactos sociales

El modelo de datos orientado a grafos logrará recomendar libros cuyo uso en las bibliotecas públicas promocionará los documentos bibliográficos invisibles o desconocidos por usuarios y por la sociedad en general.

iv. Impactos ambientales

Al ser más eficiente el modelo de base de datos orientado a grafos reduce el consumo de recursos computacionales, contribuyendo a la disminución del consumo energético.

XIII. Recursos necesarios (Infraestructura, equipos y principales tecnologías en uso relacionadas con la temática del proyecto, señale medios y recursos para realizar el proyecto)

Sistema Operativo GNU/Linux CentOS 290 • Sistema Operativo Windows 291 • Vmware 292 • Git y GitHub 293 • Flask con Python 3 294 • BDOG Neo4j

XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)

El proyecto se llevará a cabo en el distrito de Puno, provincia de Puno

XV. Cronograma de actividades

Actividad	Trimestres					
	M1	M2	M3	M4		
Analizar la arquitectura de la base de datos orientada a grafos	X					
Incorporar datos de libros sobre una base de datos orientada a grafos	X	X				
Desarrollar consultas de recomendación (filtrado híbrido)		X				
Desarrollar una interfaz de programación de aplicaciones			X			
Ejecutar pruebas			X			
Contrastar la hipótesis y formular las conclusiones				X		
Elaborar el informe final de tesis				X		
Presentar el informe final de tesis				X		

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/)	Cantidad	Costo total (S/)
Investigador	Unidad	4000.00	1	4000.00
Ordenador Portátil 4RAM 8VCORES	Unidad	2200.00	1	2200.00
Alquiler de Cloud Computing	Mensualidad	95.00	12	1080.00
Imprevistos	Unidad	500.00	1	500.00
			TOTAL	7780.00