

Universidad del Valle de Guatemala  
Fecha: Jueves, 05 de Mayo de 2021  
Algoritmos y Estructuras de datos  
Sección 20  
Docente: Melvin García

## **Proyecto 2- Algoritmos de recomendaciones**

Integrantes del grupo  
Carné 20725 Estefanía Elvira  
Carné:20673 Walter Cruz  
Carné:20289 Daniel Cabrera

## Algoritmos de recomendación

Al diseñar un motor de recomendación en tiempo real es necesario considerar no solo la información histórica del usuario, sino la sesión actual, correlacionando además al usuario con los productos, el contexto, intereses y forma en que interactúa. En este caso, y precisamente cuando nos centramos en recomendar en base a la forma en que las acciones se relacionan entre sí, es especialmente interesante evaluar la posibilidad de construir el motor empleando un grafo.

Los grafos pueden almacenarse en una base de datos donde persistirán y podrán ser consultados eficazmente. Existen numerosas bases de datos orientadas a grafos en el mercado (BDOG), como OrientDB, AllegroGraph y Neo4j, entre muchas otras.

En estos casos, lo primero es diseñar nuestro grafo conforme a cómo va a ser empleado y a los modelos que se pretenden recoger en él, lo que mejorará la eficiencia de los algoritmos que se apliquen sobre él posteriormente.

Después, conforme a ese diseño, se van generando nodos y aristas progresivamente y en tiempo real, cuando sucedan las acciones que deban persistir.

Cuando sea necesario podrán realizarse post procesamientos que añaden nuevos atributos a los nodos o a las relaciones, o incluso que generen nuevas entidades, como podrían ser nuevas aristas que representan la afinidad entre usuarios y productos, o la similitud de los usuarios entre sí.

Como ejemplo sencillo, imaginemos que tenemos un conjunto de usuarios que, a medida que realizan ciertas actividades, las valoran con puntuaciones de 1 a 10.

Podríamos determinar qué usuarios son parecidos entre sí analizando si puntuaron de manera similar las mismas actividades. Para ello, bastaría con consultar las aristas de tipo “LE\_GUSTA” de cada usuario, obteniendo el vector multidimensional del usuario, y calcular la distancia euclídea entre los vectores de cada usuarios.

Si la distancia está por debajo de cierto umbral, consideraremos que los usuarios son parecidos y creamos una nueva arista entre ambos, llamada “SIMILARITY”, con un valor conforme a la distancia. Gracias a esto podemos ordenar los usuarios por similaridad

Finalmente, se diseñan y preparan las consultas adecuadas para el tipo de recomendación que desea realizarse. Algunas de las más habituales son las siguientes:

- Productos vistos por otros usuarios: recomendar al usuario productos que no ha visto, y que otros usuarios vieron en una sesión parecida a la suya. Esta recomendación acelera el acceso a productos que previsiblemente el usuario va a consultar en la sesión, mejorando la experiencia de usuario.
- Comprados por otros: recomendar al usuario productos que otros usuarios acabaron comprando tras tener una sesión similar a la suya. Esta recomendación permite al usuario acceder rápidamente al producto que previsiblemente acabará comprando después de una fase de exploración del catálogo, reduce la posibilidad de que el usuario explore otras opciones, pero mejora la conversión en sesiones cortas o cuando el usuario no tiene mucho tiempo para decidir.
- Basado en el perfil de otros usuarios: se recomiendan productos que consumieron otros usuarios con un perfil parecido. Esta técnica permite recomendar y dar a conocer

productos a un usuario antes incluso de que empiece a explorar el catálogo, ideal para el arranque de sesión o para realizar en diferido (emailing, notificaciones push, etc.)

- Productos similares: se recomiendan productos similares a uno concreto. Esta recomendación es oportuna para ampliar la exploración del usuario, permitiéndole conocer alternativas más adecuadas a los comprados habitualmente.

Al trabajar con un grafo, si se ha diseñado correctamente, es posible construir un gran número de tipos de recomendaciones distintas sin realizar cambios. De esta forma, se dota al sistema de una colección de opciones para usar la recomendación más adecuada en cada momento o situación, optimizando los resultados de conversión y la experiencia de usuario.

## Neo4j

Neo4j es una base de datos de gráficos nativa, construida desde cero para aprovechar no solo los datos, sino también las relaciones de datos. Neo4j conecta los datos a medida que se almacenan, lo que permite consultas nunca antes imaginadas, a velocidades que nunca se creyeron posibles

### Cómo funciona Neo4j y cuáles son sus ventajas

Neo4j usa grafos para representar datos y las relaciones entre ellos. Un grafo se define como cualquier representación gráfica formada por vértices (se ilustran mediante círculos) y aristas (se muestran mediante líneas de intersección). Dentro de estas representaciones gráficas, tenemos varios tipos de grafos:

- Grafos no dirigidos: los nodos y las relaciones son intercambiables, su relación se puede interpretar en cualquier sentido. Las relaciones de amistad en la red social Facebook, por ejemplo, son de este tipo.
- Grafos dirigidos: los nodos y la relaciones no son bidireccionales por defecto. Las relaciones en Twitter son de este tipo. Un usuario puede seguir a determinados perfiles en esta red social sin que ellos le sigan a él.
- Grafos con peso: en este tipo de gráficas las relaciones entre nodos tienen algún tipo de valoración numérica. Eso permite luego hacer operaciones.
- Grafos con etiquetas: estos grafos llevan incorporadas etiquetas que pueden definir los distintos vértices y también las relaciones entre ellos. En Facebook podríamos tener nodos definidos por términos como ‘amigo’ o ‘compañero de trabajo’ y la relaciones como ‘amigo de’ o ‘socio de’.
- Grafos de propiedad: es un grafo con peso, con etiquetas y donde podemos asignar propiedades tanto a nodos como relaciones (por ejemplo, cuestiones como nombre, edad, país de residencia, nacimiento). Es el más complejo.

Neo4j utiliza grafos de propiedad para extraer valor añadido de los datos de cualquier empresa con gran rendimiento y de una forma ágil, flexible y escalable.

#### 1.- Rendimiento:

Las bases de datos orientadas a grafos como Neo4j tienen mejor rendimiento que las relacional (SQL) y las no relacional (NoSQL). La clave es que, aunque las consultas de datos aumenten exponencialmente, el rendimiento de Neo4j no desciende, frente a lo que sí sucede con las BD relacionales como MySQL.

Las BDOG responden a las consultas actualizando el nodo y la relaciones de esa búsqueda y no todo el grafo completo. Eso optimiza mucho el proceso.

Volker Pacher, desarrollador de eBay y cliente de Neo4j, explica con datos lo que supuso el cambio de MySQL a esta BDOG en el rendimiento de Shutl, la plataforma que coordina la entregas entre tiendas, mensajerías y compradores en eBay Now: “Nuestra solución Neo4j es literalmente mil veces más rápida que la solución anterior MySQL, con búsquedas que requieren entre 10 y 100 veces menos código”.

### 2.- Agilidad:

Neo4J tiene muchas ventajas, pero una es su agilidad en la gestión de datos. Si nosotros quisiéramos llevar al límite sus capacidades, tendríamos que superar un volumen total de 34.000 millones de nodos (datos), 34.000 millones de relaciones entre esos datos, 68.000 millones de propiedades y 32.000 tipos de relaciones.

### 3.- Flexibilidad y escalabilidad:

Cuando los desarrolladores de una empresa trabajan con grandes datos, buscan flexibilidad y escalabilidad. Las bases de datos orientadas a grafos aportan mucho en este sentido porque cuando aumentan las necesidades, las posibilidades de añadir más nodos y relaciones a un grafo ya existente son enormes.

## Casos de uso de Neo4j

Neo4j tiene distintos usos en la actualidad en empresas internacionales. Neo Technology tiene varios libros blancos analizando cada uno de estos usos:

### – Detección del fraude:

Neo4j ya trabaja con varias corporaciones en la detección del fraude en sectores como la banca, los seguros o el comercio electrónico. Esta base de datos puede descubrir patrones que con otro tipo de BD sería difícil de detectar.

Las redes de fraude tienen mecanismos para delinuir que no son detectables con el análisis lineal de los datos. Pero con un análisis escalable de las múltiples relaciones entre los datos, esto es mucho más fácil.

Un fraude habitual es la apertura de líneas de crédito con identidades falsas con la idea no pagar: en la actualidad, entre el 10% y el 20% de la deuda sin respaldo en los bancos líderes tanto en EEUU como en Europa se debe a este fraude.

### – Recomendaciones en tiempo real y redes sociales:

Neo4j permite conectar de forma eficaz a las personas con nuestros productos y servicios, en función de la información personal, sus perfiles en redes sociales y su actividad online reciente. En este sentido, las bases de datos orientadas a grafos son interesantes porque son capaces de conectar personas e intereses.

Con esa información, una empresa puede ajustar sus productos y servicios a su público objetivo y personalizar las recomendación en función de los perfiles. Eso es lo que permite que se aumente la precisión comercial y el compromiso del cliente.

### – Gestión de centros de datos:

Las bases de datos gráficas son el antídoto perfecto ante el crecimiento desbordante de los datos. La gran cantidad de información, dispositivos y usuarios hacen que las tecnologías tradicionales no puedan gestionar tantos datos. La flexibilidad, rendimiento y escalabilidad de

Neo4j permite gestionar, monitorizar y optimizar todo tipo de redes físicas y virtuales pese a la gran cantidad de datos.

-Gestión de sistemas de datos maestros:

La gestión de datos maestros (Master Data Management) es un auténtico quebradero de cabeza en las empresas. La creación de un sistema de información centralizado y fiable siempre es una cuestión compleja. El objetivo final es que cada miembro de una organización use los mismos formatos y aplicaciones para los datos. Eso genera un protocolo de trabajo que es aprovechable por el resto.

Neo4j ayuda a crear sistemas de ese tipo con velocidad, agilidad, rendimiento y todo eso sin perder flexibilidad y escalabilidad con los datos. Tendríamos un sistema de creación de insights en 360º: empleados, clientes y productos.

## Entrevista

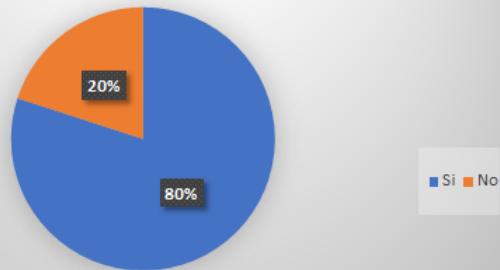
1. ¿Le resulta abrumador la cantidad y variedad de películas que tienen ahora las plataformas?
2. ¿Le gusta disfrutar ver películas?
3. ¿Cree que es bueno dedicarle tiempo al cine y a las producciones audiovisuales de vez en cuando?
4. ¿Cuánto tiempo tarda para elegir una película cuando quiere ver algo su plataforma de preferencia?
5. ¿Cuánto tiempo considera prudente para elegir una película que ver?
6. ¿Le gustaría poder ahorrar tiempo para poder elegir?
7. ¿Usted considera una buena propuesta el usar una herramienta que le ayude a decidir más rápido?

## Gráficas

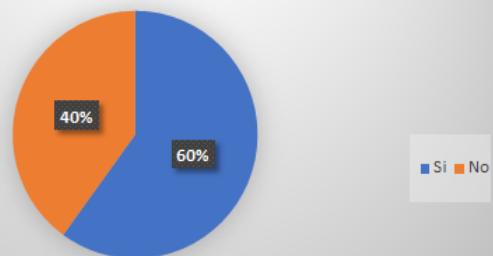
¿Le resulta abrumador la cantidad y variedad de películas que tienen ahora las plataformas?



¿Le gusta disfrutar ver películas?

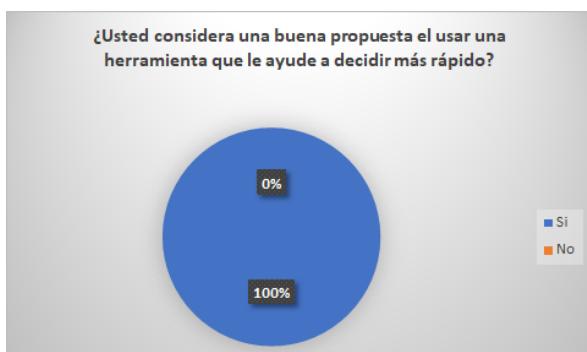


¿Cree que es bueno dedicarle tiempo al cine y a las producciones audiovisuales de vez en cuando?



¿Cuánto tiempo tarda para elegir una película cuando quiere ver algo su plataforma de preferencia?





## DEFINICIÓN DEL PROBLEMA

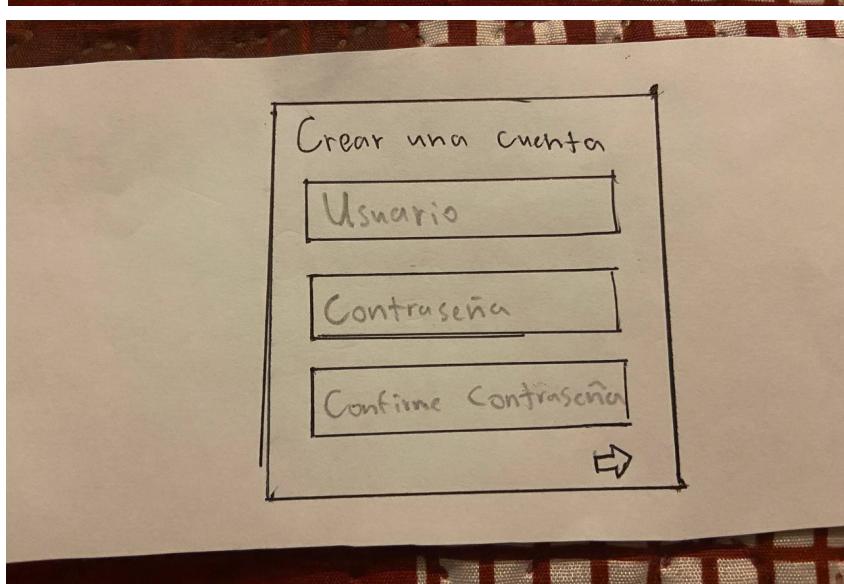
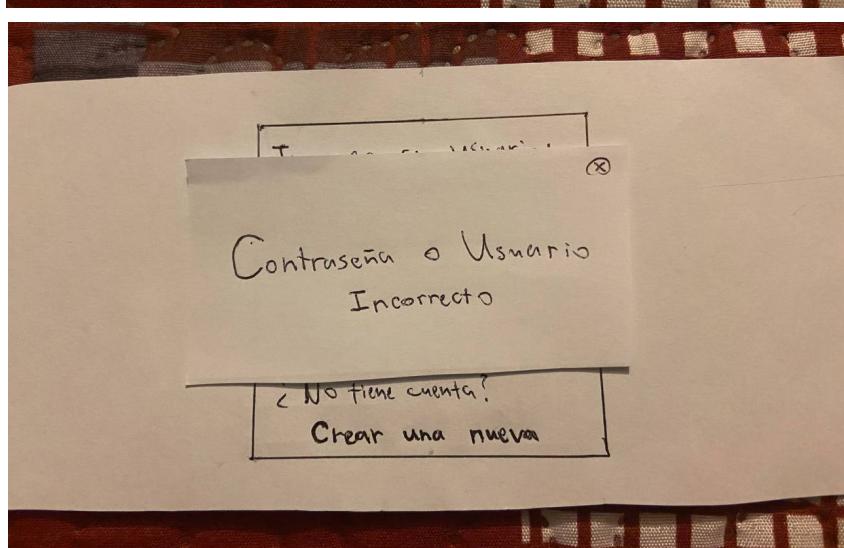
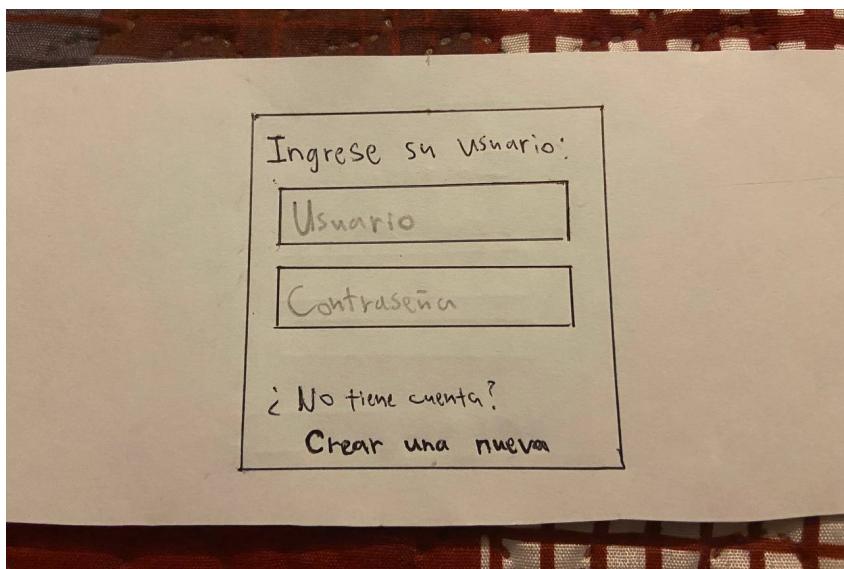
La problemática que se encontró para la realización de este proyecto, es que a las personas se les dificulta elegir una película que sea acorde a mis preferencias entre el catálogo tan grande que tiene las plataformas de streaming hoy en día, pues en ciertas ocasiones no encuentran nada que les llame la atención.

## IDEACIÓN

- Hacer un programa que permita ayudar a los usuarios a escoger una película de una base de datos basándose en los géneros más vistos por el usuario y por popularidad con los demás.
- Hacer un programa que ayude a los usuarios a escoger una película recomendando por su tiempo de duración y el tiempo de duración de las películas que ha visto.
- Hacer un programa que recomiendan películas de forma aleatoria en base a las preferencias de tus amigos más cercanos.
- Hacer un programa con una enorme lista de películas con una especie de top y que estas se vayan actualizando según la popularidad de las mismas.

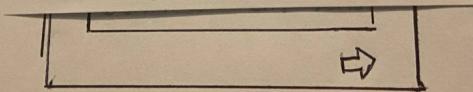
## PROTOTIPO

Mock-up de papel:



### Crear una cuenta

Contraseña no coincide.  
Intente de nuevo.



### ¡Bienvenido!

Indique cuales son sus géneros favoritos:

Terror

Ciencia Ficción

Acción

Suspense

Románticos

Comedia

Otros..

### Siguiente

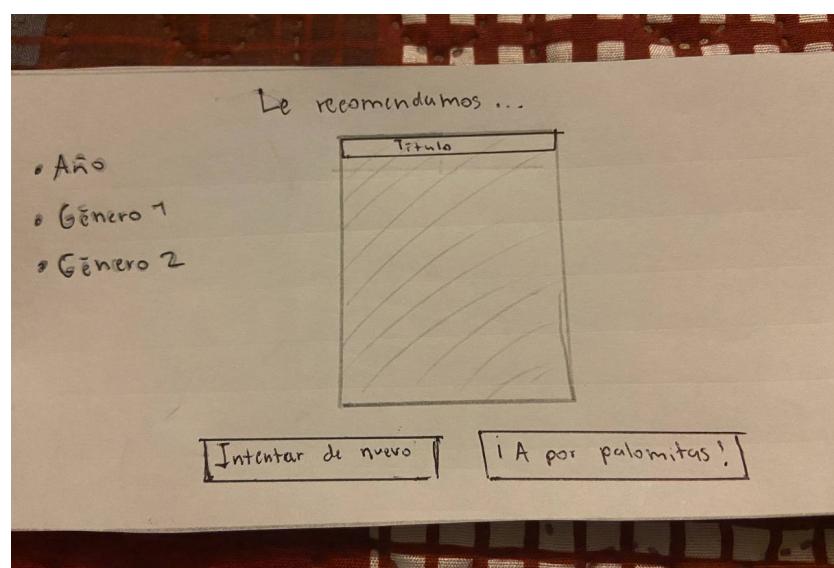
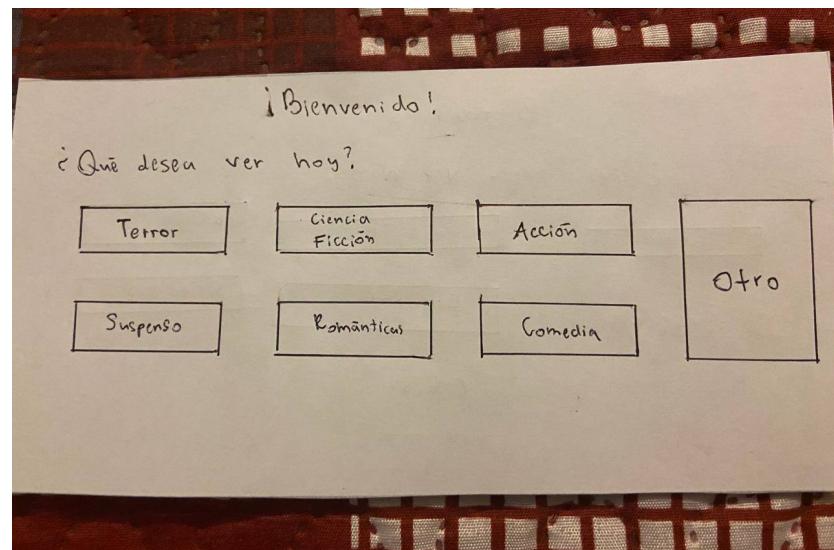
Por favor ingrese 3 de sus películas favoritas...

Película 1

Película 2

Película 3

### Siguiente



## REFERENCIAS

- <https://www.solusoft.es/innovacion/sistemas-recomendacion-basados-en-grafos>
- <https://neo4j.com/developer/graph-database/>
- <https://www.bbva.com/es/que-es-neo4j-y-para-que-sirve-una-base-de-datos-orientada-a-grafos/>
- <https://neo4j.com/developer/language-guides/>