

Algoritmos y Estructura de Datos
2019
Proyecto 3
Enriquez
Grupo #8
Argüello
Integrantes:
Michael Chan
Diego Estrada 18540
Isabel Ortiz

Guatemala 3 de mayo de

Catedrático: Diego

Auxiliares: Andrea

Ana Lucía Hernández



Proyecto 3

Fase 1

Design Thinking

Investigación sobre motores de recomendación:

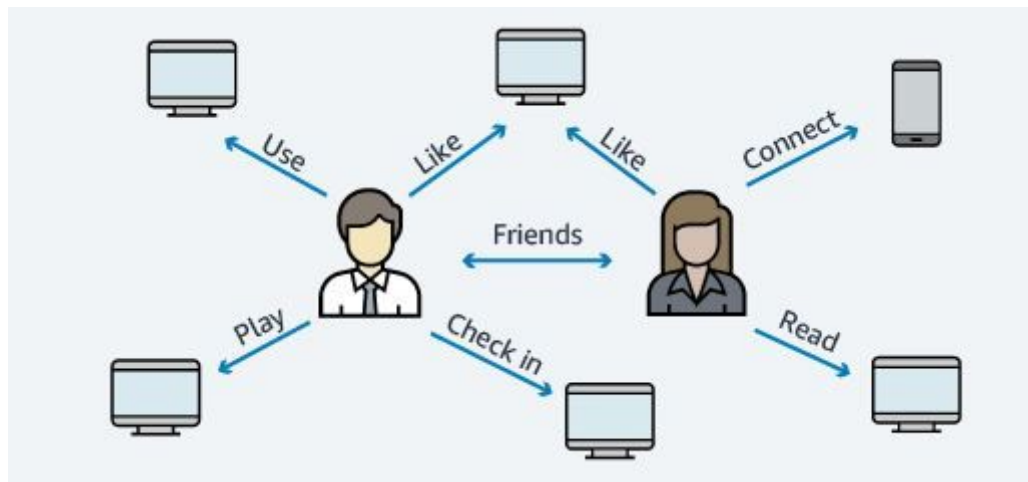
Un motor de recomendación o *recommendation engine* es un conjunto de herramientas de software que se encarga de proveer al usuario que lo utiliza una serie de recomendaciones o sugerencias con respecto a un producto o servicio, basándose en las cualidades personales del propio usuario. El algoritmo del motor se encarga de realizar un análisis basándose en las preferencias del usuario, toma de decisiones y cualidades del consumidor. Este tipo de software es aplicado en muchos ámbitos, uno de ellos es el comercio en línea o *e-commerce*, en donde conocer exhaustivamente al usuario es sumamente importante para prolongar su consumo o fidelidad a la marca. Ésta es una de las formas para promover la competitividad en el mercado, tanto de compras en línea, entretenimiento e incluso comida.

Los motores de recomendación van desde algo tan simple como recomendaciones de familiares y/o amigos para comprar ropa, hasta programas desarrollados como *LinkedIn* para encontrar trabajo. Muchos de los motores en línea están basados en grafos, los cuales son representaciones gráficas de objetos conocidos como nodos, que a su vez están conectados por aristas, ellos permiten establecer relaciones de carácter binario entre los objetos de un conglomerado. Dentro de los motores basados en grafos se puede mencionar: Facebook: esta red social cuenta con *People You May Know*, el cual recomienda usuarios basándose en amigos en común, de manera que es posible que conozca la persona fuera de internet; Netflix: para este servicio de streaming, la función *Other Movies You Might Enjoy* buscará películas para recomendar tomando como referencia el historial de filmes que el usuario ha visto a través de su navegación, tomando en cuenta factores como:

género y calificaciones previas hechas por el usuario;

LinkedIn: con *Jobs You May be Interested In* se recomienda tanto a la empresa como al usuario basado en los requerimientos y habilidades del perfil, así como también la experiencia previa.

Otro ejemplo interesante de motores de recomendación basados en grafos es *Amazon Neptune*. Este es un servicio de base de datos que permite almacenar millardos de relaciones y así mismo consultar grafos con una gran rapidez. Con *Amazon Neptune* se permite al usuario adaptar cualquier aplicación tales como redes sociales a un motor de recomendación avanzado, permitiendo sugerir a los consumidores diversos productos que pueden ser de su agrado basándose en consumos anteriores. En él se admiten grafos de tipo *W3C's RDF* y *Property Graph*, cada uno junto a sus lenguajes de consulta correspondientes, como *Apache TinkerPop Gremlin* and *SPARQL*; además de motor de recomendación, *Neptune* permite implementar detectores de fraude y seguridad de redes.



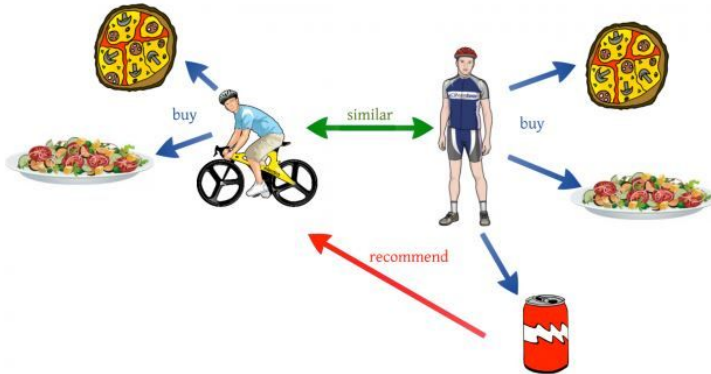
tomado de: https://aws.amazon.com/neptune/?nc1=h_ls

Es posible clasificar los motores de recomendación en varios tipos según su algoritmo, de los cuales se puede mencionar: Filtración colaborativa, Filtración basada en contenido y sistemas híbridos.

Filtración Colaborativa:

Los métodos de filtración de este tipo recaban y analizan una gran cantidad de información sobre los usuarios, desde comportamientos, preferencias y actividades habituales, para finalmente realizar predicciones sobre qué le gustaría al usuario usando las similitudes con otros usuarios. Este tipo de filtración tiene la ventaja de que prescinde del entendimiento del contenido analizable, es decir, que es capaz de

recomendar objetos complejos sin la necesidad de comprender el elemento en sí (Pepa, 2014).



tomado de: <https://dataconomy.com/2015/03/an-introduction-to-recommendation-engines/>

El fundamento de este paradigma radica en que el siguiente caso se agrupará en la clase con mayor frecuencia a la que sus K vecinos más cercanos pertenecen.

		X_1	...	X_j	...	X_n	C
(x_1, c_1)	1	x_{11}	...	x_{1j}	...	x_{1n}	c_1
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
(x_i, c_i)	i	x_{i1}	...	x_{ij}	...	x_{in}	c_i
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
(x_N, c_N)	N	x_{N1}	...	x_{Nj}	...	x_{Nn}	c_N
x	$N + 1$	$x_{N+1,1}$...	$x_{N+1,j}$...	$x_{N+1,n}$?

Figura 1: Notación para el paradigma K-NN

Se muestra la figura anterior y se observa la indicación de N casos, en donde cada uno está representado por n variables predictoras, y finalmente una clase C a ser determinada.

COMIENZO

Entrada: $D = \{(x_1, c_1), \dots, (x_N, c_N)\}$

$x = (x_1, \dots, x_n)$ nuevo caso a clasificar

PARA todo objeto ya clasificado (x_i, c_i)

calcular $d_i = d(x_i, x)$

Ordenar $d_i (i = 1, \dots, N)$ en orden ascendente

Quedarnos con los K casos D_x^K ya clasificados más cercanos a x

Asignar a x la clase más frecuente en D_x^K

FIN

(Moujahid, Inza, &

Larrañaga, 2014)

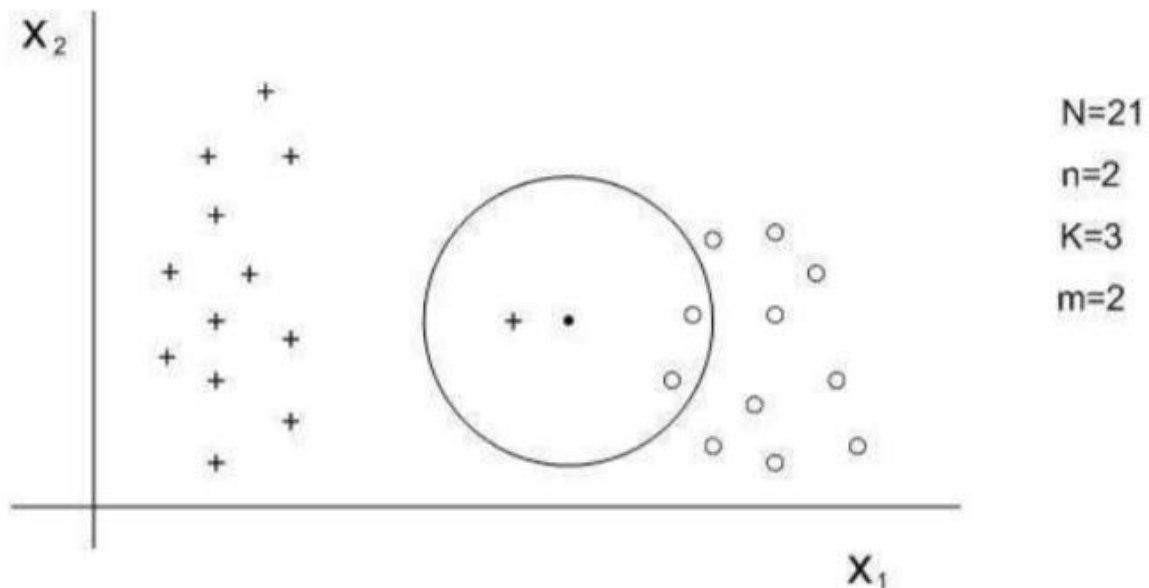


Figura 3: Ejemplo de aplicación del algoritmo K-NN básico

(Moujahid, Inza, & Larrañaga, 2014)

En la figura 3 se representan 24 casos clasificados con una cantidad de posibles valores $m = 2$. Compuesta por variables predictoras X_1 y X_2 y valor de K de 3. Es posible observar que 3 casos han sido ya clasificados y se encuentran próximos al nuevo caso representado por un punto relleno, y por lo tanto el clasificador 3-NN predice su clase.

Las variantes que existen para el clasificador K-NN son: con rechazo y con distancia media. La primera de ellas tienen el fundamento de que con el fin de realizar una clasificación para dado caso, es necesario poseer garantías. Esto abre espacio a que el caso no sea clasificado ya que no existen las garantías requeridas para que la asignación de clase sea la adecuada. La segunda variante, de distancia media, consiste en la asignación de un caso a una clase cuya distancia media sea inferior. En la figura 5 se observa que, a pesar de 5/7 casos pertenecen a una misma clase, el caso nuevo se clasificará como el más cercano, dependiendo de la distancia media y no de la frecuencia.

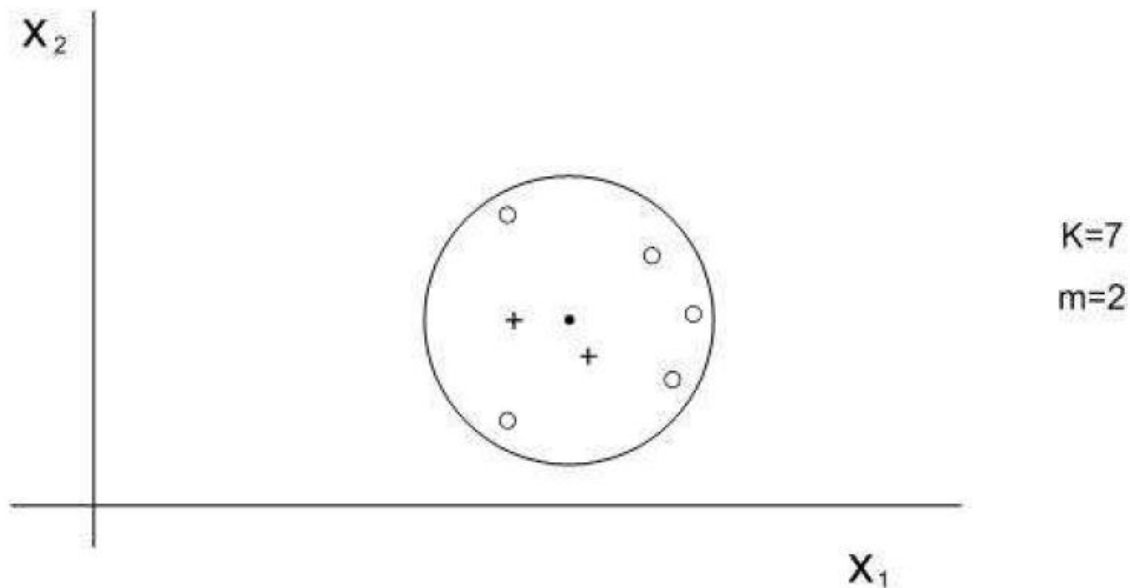
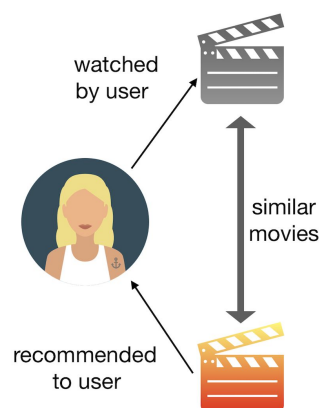


Figura 5: Ejemplo de ilustración del K-NN con distancia media

(Moujahid, Inza, & Larrañaga, 2014)

Filtración Basada en Contenido:

Para esta filtración, los métodos toman como base la descripción y características de los productos de preferencia del usuario. Se caracteriza por la utilización de palabras clave para crear las relaciones, además, el perfil de usuario se crea con el fin de indicar sus gustos personales. Estos algoritmos pretenden realizar sugerencias basándose en ítems que en el pasado han sido del agrado del consumidor. Dentro de todas las posibles sugerencias, el algoritmo tomará aquellas mejores posicionadas y calificadas por el usuario, de esta manera la selección será más útil.



fuelle:

<https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243>

Dentro del algoritmo de filtración basada en contenido, se encuentra la variante Rocchio, la cual consiste en calcular los centroides para cada usuario y así obtener un vector característico para cada uno. Se sigue implementando la lógica de considerar el contenido del ítem, tal como facebook con el contenido de los *posts*, de esta manera cada usuario obtiene un centroide relacionado a la característica del ítem.

Su representación matemática es la siguiente:

$$f(u, i) = \text{sim}(u, i) = \frac{\sum_f u[f] * tfidf(f, i)}{\sqrt{\sum_f u[f]^2} \sqrt{\sum_f tfidf(f, i)^2}}$$

Por último, el algoritmo llamado ítem kNN, tienen similitudes con su homónimo en la filtración colaborativa, a diferencia de que en este algoritmo se toma en cuenta únicamente la descripción y características de los ítems.

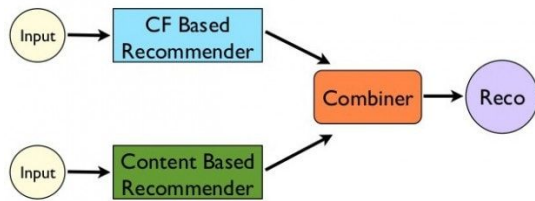
Su representación matemática es la siguiente:

$$\text{sim}(i, j) = \frac{\sum_f tfidf(f, i) * tfidf(f, j)}{\sqrt{\sum_f tfidf(f, i)^2} \sqrt{\sum_f tfidf(f, j)^2}}$$

Recomendaciones Híbridas:

Como su nombre lo intuye, este tipo incluye características de las filtraciones colaborativas y basadas en contenido, y que según investigaciones, puede ser más efectivo que las dos individuales. Pueden ser aplicadas de distintas maneras, la colaborativa separada de la basada en contenido y después combinarlas, o simplemente unificarlas desde un principio. El método híbrido es capaz de ofrecer recomendaciones más precisas que las filtraciones puras por separado, ya que existen elementos que se olvidan y no se toman en cuenta en un método que podrían ser útiles en otro. Un ejemplo de este motor de recomendación es *Netflix*, quien realiza recomendaciones basándose en lo que mira el usuario en comparación a lo que busca, y así mismo, toma en cuenta las características de las películas que ha visto, todo esto para ofrecerle una sugerencia con alta precisión al usuario.

Hybrid Recommendations



tomado de: <https://dataconomy.com/2015/03/an-introduction-to-recommendation-engines/>

Fase Empatía:

- **Definición del problema:** Como grupo se encontró que actualmente no hay un sitio de internet confiable, donde se hagan las preguntas pertinentes y se averigüe lo necesario, para encontrar a la pareja ideal. Pues se ha escuchado de muchas personas que van por estos sitios, y que no han tenido experiencias favorables. Pues pasan por situaciones vergonzosas o simplemente les toca vivir un mal rato.
- **Propuesta de ideas de solución del problema:** Para mejorar la problemática anterior, se propuso hacer un sistema de recomendación de citas, donde la información será segura, y también se hagan preguntas coherentes, tocando temas de gustos, pasiones, y preferencias. Para que pueda filtrarse la información y así sea más efectivo el “match”.
- **Entrevistas:** Para poder tener información más útil respecto a este tema, se seleccionó a un grupo de 8 personas que tienen entre 19 y 24 años, que han hecho uso o aún utilizan la app “Tinder”, que permite a los usuarios comunicarse con otras personas con base en sus preferencias para charlar y concretar citas o encuentros. Se hicieron preguntas respecto al uso de aplicación, qué les gustaría que les preguntara, qué no les gusta de la aplicación en general, y recomendaciones para nuestra página. La información dada en estas entrevistas, hizo que se pudiera trabajar con más claridad respecto a la realización de la página, pues se supo específicamente que quería encontrar el usuario, especialmente lo útil y beneficioso.

Perfiles:

- Jóvenes entre un rango de 18 y 22 años que hagan uso de Tinder (aplicación de recomendación de citas).

Las entrevistas tenían las siguientes preguntas:

1. Qué tipo de preguntas te gustaría que te hicieran en un sitio de recomendación de citas?
2. Haciendo referencia a la aplicación móvil Tinder, qué no te gusta de las preguntas que se hacen o de la modalidad de la aplicación?
3. Cómo fue tu experiencia utilizando Tinder?
4. Cómo sería el sitio de recomendación de citas ideal para ti? (Preguntas, información personal, etc.)

Perfiles encuestados:

**Oscar López, estudiante de medicina de la Universidad de Guadalajara
20 años**



“Me daba cuenta que mis amigos usaban Tinder, y por curiosidad lo descargué”

Opinión: él compartió que la app sería muy útil para personas que buscan una relación formal, pues tomará en cuenta aspectos personales y no solo lo físico. Considera también importante las recomendaciones basadas en la ubicación de la otra persona. Ha tenido buenas experiencias y lo ha usado numerosas ocasiones. Su programa de citas ideal consultaría cosas más personales y de gustos, de manera que el encuentro sea mejor y que la relación funcione.

**Peter Pratdesaba
Estudiante de ingeniería de la Universidad de San Carlos de Guatemala
18 años**



“Instale la aplicación solamente para conocer chavas”

Opinión: él piensa que la aplicación sería bastante útil especialmente para personas que son tímidas y les cuesta tomar iniciativa respecto a las relaciones amorosas.

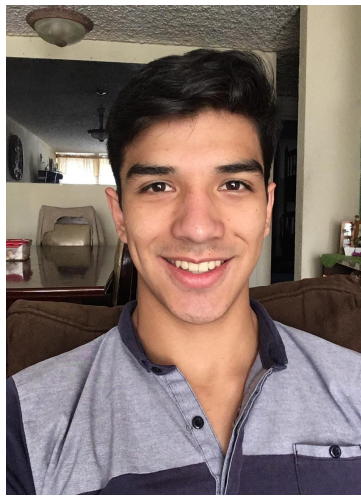
Mercy Peters,
estudiante de Relaciones Internacionales en la Universidad de San Carlos
de Guatemala
18 años



“Yo uso tinder regularmente, aunque no me gusta del todo porque he tenido malas experiencias”

Opinión: Ella considera que lo ideal sería preguntarle al usuario sus preferencias musicales, hobbies, comida, etc. Tal y como lo hace tinder, con la diferencia de que ha tenido citas en las que la persona no concuerda con el perfil que había establecido.

Diego Álvarez
Estudiante de ingeniería de la Universidad de San Carlos de Guatemala
18 años



“Descargué Tinder porque me dió curiosidad, pero me llevé varias sorpresas”

Opinión: él piensa que es importante tomar en cuenta la personalidad y gustos de las personas, para poder hacer un buen “match” entre personas.

Manuel Cruz
Estudiante de Artes Culinarias en el Intecap
21 años



“Descargué Tinder porque acababa de terminar con mi novia, y quería algo pasajero”

Opinión: él piensa que la idea de la aplicación es buena, pues los jóvenes hoy en día tienen más confianza de hablar con alguna persona de interés por medio del celular. Menciona que le gustaría un *match* basado más en gustos similares como *hobbies* que en una foto. Prefiere recomendaciones más basadas en la personalidad del usuario.

Juan David Cruz
Estudiante de licenciatura en música de la Universidad Galileo
22 años



“Descargué Tinder por una apuesta, pero encontré a varias chicas que llamaron mi atención ”

Opinión: él piensa que debemos hacer preguntas bastante personales, para poder tener un idea concreta de cómo es realmente la persona. Opina que es una buena herramienta para conocer personas, pero que antes de la cita no se llega a conocer a fondo a la persona, pues solo tiene una foto y una breve descripción.

Emidio Corado
Estudiante de ingeniería industrial de la Universidad de San Carlos
20 años



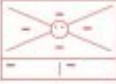
“Descargar tinder me permitió conocer a mi novia actual ”

Opinión: él piensa que debemos preguntar gustos musicales, literarios, de series o películas, de comida, y distintos temas. Para así tener bastantes campos para poder hacer las distintas elecciones. Le cuesta encontrar una buena coincidencia porque la app no pregunta cosas muy específicas, todo depende de la biografía y no de grupos con gustos similares.

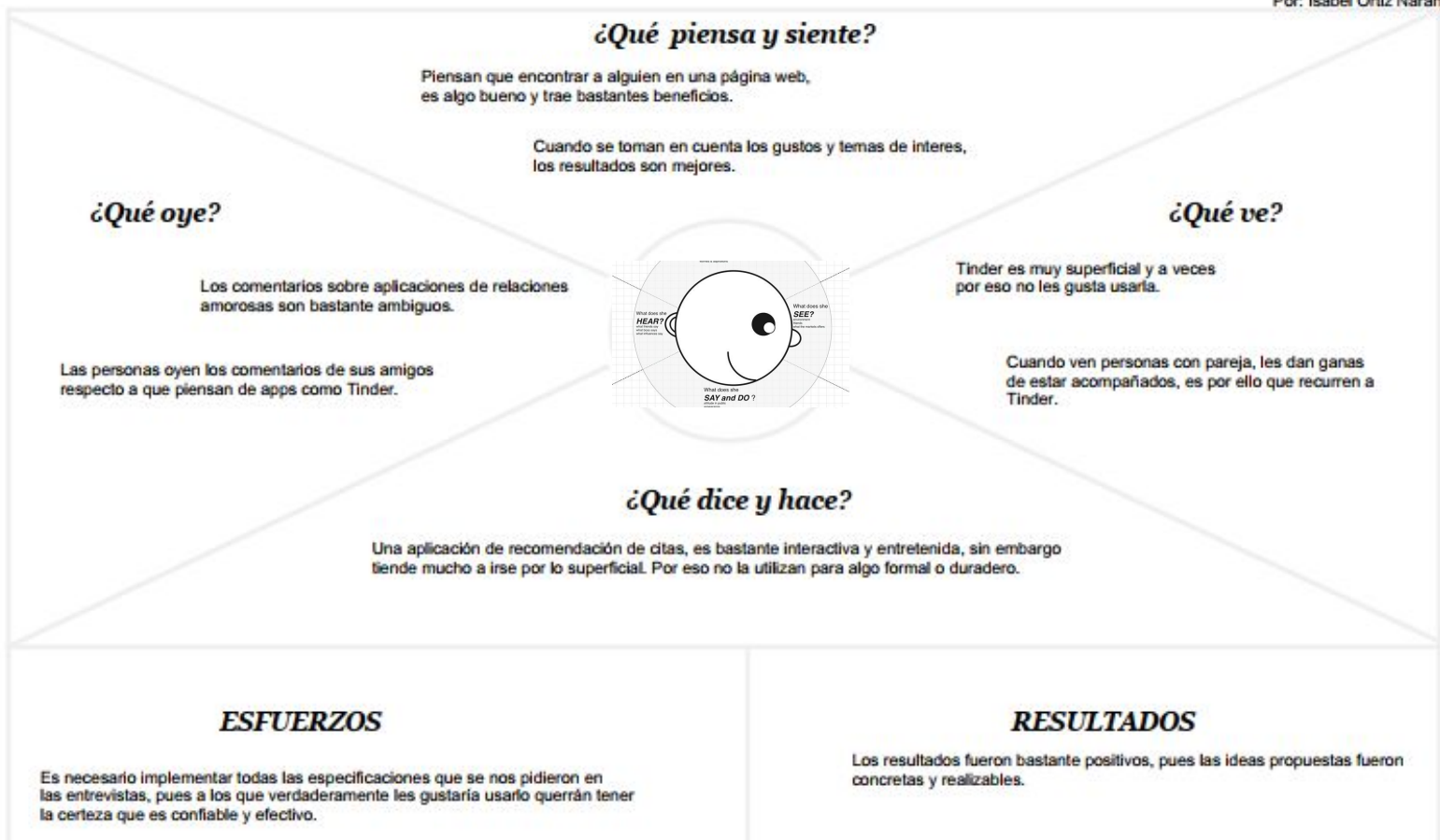
Mapa de Empatía:

Mapa de empatía

Nombre del usuario/cliente:
Jóvenes en el rango de edad de los 18 a 22 años.



Por: Isabel Ortiz Naranjo



Necesidades:

- Actualmente las personas que desean encontrar pareja en alguna página web, tienen miedo que la información que se les proporciona sea falsa o bien la información que estos dan se utilizada para otros fines.
- Tener una relación permite tener más confianza con uno mismo.
- Que las personas que son solitarias o tímidas, se den la oportunidad de conocer personas con sus mismos gustos y así puedan llevar una relación bastante enriquecedora.

Insights:

- Hacer match entre personas con los mismos gustos.
- Hacer las relaciones adecuadas, según la ubicación de cada persona.
- Mostrar la funcionalidad de la aplicación, según las coincidencias de cada uno de los usuarios y por estas similitudes, realizar el match correspondiente.

Fase de Ideación:

Problema:

Ausencia de un sistema de citas confiable, donde se hagan las preguntas pertinentes y se averigüe lo necesario, para encontrar a la pareja ideal y establecer una relación íntima. Se considera un problema pues se ha escuchado de muchas personas que van por estos sitios, y que no han tenido experiencias favorables, pues pasan por situaciones vergonzosas o simplemente les toca vivir un mal rato debido a que la persona o no cumple con el perfil, o bien hubo detalles que no se mencionaron en dicha plataforma.

Lluvia de ideas:

- Organizar a las personas por “nichos”, de tal manera que se relacionan entre ellas compartiendo gustos.
- Basar las relaciones puramente en los hobbies de los usuarios
- Tomar en cuenta los rangos de edad y recomendar según las preferencias para esta clasificación
- Relaciona principalmente y dar prioridad a la ciudad en la que vive el usuario, de manera que el acceso a una cita es más fácil.
- Un sitio de citas a ciegas en donde solo la personalidad de la persona importe, se desplegará un listado de cualidades intrínsecas y el usuario las elegirá, con base en sus elecciones tendrá un puntaje. Con dicho puntaje se realizará el *match* utilizando los rangos de personalidad.

Idea Final y justificación:

La idea que se tomará para realizar, en primera instancia el prototipo, y posteriormente el programa final, será la de una cita basada más en la personalidad de las personas, donde la foto y rasgos físicos sea lo que importe menos. Se le preguntará al usuario su tipo de comida favorita, género de música y de película que más le gusta, ubicación y dónde vive, todo esto para realizar un *match* final, en donde se le muestre al usuario, ordenado por número de coincidencias, todas las personas con las que sería buena idea salir, basándose en las similitudes.

Fase Testing:

- Para esta fase, se hicieron pruebas de acuerdo al prototipo realizado. Los usuarios hicieron retroalimentaciones según lo que ellos consideraban como mejora, y el resultado del prototipo es como está a continuación:
 - Se podrán observar debajo de cada una de la imagen las recomendaciones dadas.

1.

¡Bienvenido! ¿Estás listo
para tu próxima cita?

Empezar

- Se recomendó que se pusiera una frase o pregunta que motivara al usuario para ingresar sus datos y empezar con la experiencia.

2.

Información Personal:

¿Como te llamas?

Tu nombre

¿Cuántos años tienes?

Tu edad

¿Dónde vives?(Ciudad, País)

Ubicación

Siguiente

- Se recomendó que se agregara la ciudad, no solo el país donde habita el usuario interesado.

3.

¿Qué tipo de música
disfrutas más?

- ☐ Hip Hop
- ☐ Reggaetón
- ☐ Rock
- ☒ RnB
- ☐ Clásica
- ☐ Cumbia
- ☐ Electrónica

Siguiente

- Se recomendó que hubiera más opciones de música, por ello se seleccionaron 7 géneros.

4.

¿Cuál es tu comida
favorita?

- ☐ China
- ☐ Italiana
- ☐ Fast Food
- ☒ Chapina
- ☐ Sushi
- ☐ Mexicana

Siguiente

- Se recomendó que hubiera más opciones de comida, por ello se seleccionaron 6 tipos.

5.

¿Qué género de películas consideras tu favorito?

- ☐ Acción
- ☐ Miedo
- ☐ Tragedia
- ☒ Comedia
- ☐ Drama
- ☐ Superhéroes

Siguiente

- Un usuario recomendó agregar el género de los superhéroes, pues es el tema de los jóvenes en estos momentos.

6.

[Regresar al Inicio](#)

No sería mala idea tener una cita con alguna de estas personas ... ;)

Tu match 	
Diego Estrada	Ver
Michael Chan	Ver
Saul Contreras	Ver
Adriana Castillo	Ver

- Un usuario recomendó el poner una frase donde se dé a conocer que esa lista de personas, hacen match con el usuario que está interesado.

7.



A user profile card for Michael Chan. It features a circular placeholder for a profile picture at the top left. To the right of the profile picture is a rectangular button labeled "Regresar". Below the profile picture, the text "Nombre: Michael Chan" is displayed. Underneath the name, it says "Edad: 18 años". Then, "Comida Favorita: China". Below that, "Le gustan las películas de: Acción". Then, "Disfruta de la música: Rock". At the bottom of the card is a rectangular button labeled "Contactar".

Regresar

Nombre: Michael Chan

Edad: 18 años

Comida Favorita: China

**Le gustan las películas
de: Acción**

**Disfruta de la música:
Rock**

Contactar

- Se hizo la recomendación, que cuando se despliegan los datos del usuario, haya un botón que haga el contacto directamente.

Algoritmo narrativo →

Pseudocódigo:

-
1. Crear la base de datos a partir de los datos ingresados por los usuarios, entendiéndose su nombre, edad y ubicación.
 2. Conectar la base de datos a través de Neo4J.
 3. Se crea un arreglo con los todos usuarios ingresados a la base de datos.
 - 3.1 Se verifica que el usuario que ingrese sus datos no tenga una cuenta, de lo contrario se pondrá un mensaje diciendo que la cuenta ingresada como nueva ya existe.
 4. Se crea arreglos vacíos para las personas que coincidan en gustos musicales.
 5. Se crea arreglos vacíos para clasificar a las personas que coincidan en gustos de comida.
 6. Se crea arreglos vacíos para clasificar a las personas que coinciden en género de películas favorito.
 7. El programa da la bienvenida, y pregunta si está el usuario listo para su próxima cita.
 8. Se despliegan los enunciados preguntando al usuario su nombre, edad y ubicación, y se irá guardando esta información en el arreglo vacío de los datos del usuario.
 9. Se hace la verificación indicada en el numeral **3.1**
 10. Se despliega la pregunta sobre qué tipo de música el usuario prefiere escuchar.
 11. Se despliega todos los géneros musicales seleccionados por los creadores del programa
(nosotros).
 - 11.1 Se va haciendo el análisis de las coincidencias con los otros usuarios.
 12. Se despliega la pregunta sobre qué tipo de comida el usuario prefiere o es su favorita.
 13. Se despliega todos los tipos de comida seleccionados por los creadores del programa
(nosotros).
 - 12.1 Se va haciendo el análisis de las coincidencias con los otros usuarios.
 14. Se despliega la pregunta sobre qué tipo películas el usuario prefiere ver.
 15. Se despliega todos los géneros de películas seleccionados por los creadores del programa
(nosotros).
 - 15.1 Se va haciendo el análisis de las coincidencias con los otros usuarios.

16. Según el máximo de coincidencias que el usuario que ingresó a hacer las pruebas tenga con los demás usuarios, se le brindará una lista de todos los “match” que este hizo.
17. En la lista de los “match” se puede ver el usuario completo, con los datos completos a disposición del usuario interesado.
18. Se pregunta si el usuario desea regresar al inicio o salir del programa.
19. Salir del programa
 - 19.1 Fin del programa.

Literatura Citada:

-
- Pepa, S. M. (1 de Mayo de 2014). Suite de Algoritmos de recomendacion en aplicaciones reales. Recuperado el 4 de 30 de 2018, de https://repositorio.uam.es/bitstream/handle/10486/660903/marina_pepa_sofia_tfg.pdf?sequence=1
- Moujahid, A., Inza, I., & Larrañaga, P. (2014). Clasificadores K-NN. Recuperado el 30 de Abril de 2018, de <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t9knn.pdf>