



**BUAP**

**Benemérita Universidad Autónoma de Puebla**

**Facultad de Ciencias de la Computación**

**Sistema de recomendación de artículos científicos en el ámbito de la Medicina: Caso de estudio Cáncer de Mama – Diagnóstico, tratamiento y prevención**

**Tesis para obtener el grado de:  
Maestría en Ciencias de la Computación**

**Presenta:**  
Oscar Eduardo González Ramos

**Director de tesis:**  
Dra. María de la Concepción Pérez de Celis Herrero

**Co-Asesor de tesis:**  
Dra. María Josefa Somodevilla García



*H. Puebla de Zaragoza, Noviembre 2022*



## **Dedicatoria**

A todas las personas que han estado apoyándome en este camino, quienes de manera incondicional siempre me han alentado a dar lo mejor de mí en cada reto, haciendo de esto una experiencia de lo más preciosa.

## **Agradecimientos**

A Dios, por permitirme tener cada día una nueva oportunidad de ser mejor persona, por cada lección de vida y por cada uno de los obstáculos que me ha ayudado a superar.

A mis padres y hermanos, quienes, me han otorgado su apoyo incondicionalmente en todos y cada uno de los retos que se presentan día con día.

A Ademaris, por brindarme su apoyo y amor incondicional día tras día, por ser uno de los pilares más importantes en mi vida.

A los colegas de la maestría, quienes, con su apoyo, consejos y sobre todo su conocimiento logramos aprender en conjunto y disfrutar al mismo tiempo esta etapa.

A cada uno de los docentes, que compartieron su confianza y conocimientos durante este periodo. En especial a mi asesora, Dra. María de la Concepción Pérez de Celis Herrero, porque nunca dejo de apoyarme y siempre alentaba a continuar con el desarrollo de este trabajo.

Finalmente, no menos importante, agradecimiento especial al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado durante los estudios de postgrado.

## Contenido

Resumen .....	1
Capítulo 1 .....	2
1.1 Introducción .....	2
1.2 Motivación y Objetivos.....	3
1.2.1 Objetivo General.....	3
1.2.2 Objetivos Específicos .....	3
1.3 Metodología.....	4
Capítulo 2 .....	5
2.1 Estado del Arte .....	5
2.1.1 Antecedentes del Proyecto.....	5
2.1.2 Sistemas de Recomendación.....	12
2.1.2.1 Definición Formal.....	12
2.1.2.2 Algoritmos de filtrado.....	12
2.1.2.3 Técnicas de Recomendación .....	13
Capítulo 3 .....	16
3.1 Análisis y especificación de requisitos .....	16
3.2 Casos de Uso .....	17
3.2.1 Actores.....	17
Capítulo 4 .....	19
4.1 Diseño de la Taxonomía.....	19
4.1.1 Proceso de creación.....	23
4.1.2 Conceptos.....	24
4.1.3 Visualización de la Taxonomía implementada .....	40
Capítulo 5 .....	45
5.1 Arquitectura del Sistema de Recomendación .....	45
5.1.1 Modelo de datos .....	46
5.1.2 Rastreo y búsqueda de PDF académicos.....	51
5.1.2.1 Web Scraping.....	51
5.1.2.2 Desventajas del Web Scraping .....	53

5.1.2.3 Alternativa del web scraping.....	54
5.1.2.4 Ventajas de las API .....	54
Capítulo 6 .....	56
6.1 Implementación .....	56
6.1.1 Inicio de sesión .....	58
6.1.2 Panel Principal (Dashboard).....	58
6.1.3 Búsqueda.....	59
6.1.4 Resultados de una Búsqueda .....	60
6.1.5 Cerrar sesión .....	61
Capítulo 7 .....	62
7.1 Codificación .....	62
7.1.1 Estructura del Proyecto .....	62
Capítulo 8 .....	70
8.1 Resultados.....	70
Capítulo 9 .....	71
9.1 Conclusiones y Trabajo a futuro .....	71
Bibliografía.....	72

## Índice de Ilustraciones

Figura 1 - Caso de Uso Usuario General .....	18
Figura 2 - Clasificación medica cáncer de mama .....	19
Figura 3 - Clasificación del Cáncer de Mama .....	20
Figura 4 - Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario. ....	21
Figura 5 - Estructura general.....	22
Figura 6 - Diagrama de flujo para la creación de la Taxonomía .....	23
Figura 7 - Definición de entidades en Protégé .....	40
Figura 8 - Taxonomía horizontal .....	42
Figura 9 - Taxonomía Radial, facetas y algunos conceptos .....	43
Figura 10 - Conceptos completos de la Taxonomía .....	44
Figura 11 - Ejemplo de modelo Cliente-Servidor .....	45
Figura 12 - Protocolo de Transferencia de Archivos.....	46
Figura 13 - Esquema Conceptual de la Taxonomía Implementada .....	48
Figura 14 - Modelo Relacional del proyecto .....	49
Figura 15 - Representación del Web Scraping.....	51
Figura 16 - Pruebas Selenium.....	52
Figura 17 - Pruebas Beautiful Soup.....	53
Figura 18 - Inicio de sesión.....	58
Figura 19 - Panel Principal / Dashboard .....	59
Figura 20 - Buscador.....	59
Figura 21 - Resultados de una búsqueda .....	60
Figura 22 - Cerrar sesión .....	61
Figura 23 - Almacén de información de los recursos.....	61
Figura 24 - Arquitectura del Proyecto.....	63
Figura 25 - Librerías y Ruta Principal.....	64
Figura 26 - Login.html.....	64
Figura 27 - Ruta Dashboard.....	65
Figura 28 - Código de la Taxonomía.....	66
Figura 29 - Consultas a la Base de Datos.....	67
Figura 30 - Recuperación de Recursos Parte 1.....	68
Figura 31 - Recuperación de Recursos Parte 2.....	69
Figura 32 - Definición del método principal .....	69

## Índice de Tablas

Tabla 1 - Diccionario de Conceptos de la Taxonomía.....	39
Tabla 2 - Entidades de la BD.....	50



## Resumen

Los sistemas de recomendación son un tema fascinante tanto para los investigadores como para la industria. Los investigadores encuentran en los sistemas de recomendación un tema que es relevante para muchas disciplinas: aprendizaje automático, minería de textos, inteligencia artificial, análisis de redes, bibliometría, bases de datos, computación en la nube, escalabilidad, ciencia de datos, visualización, computadora humana interacción y muchos más, eso hace que la investigación en sistemas de recomendación sea interesante. En los sistemas de recomendación, la priorización y la personalización de la información se encuentra ausente, lo que ha incrementado la demanda de nuevos sistemas de recomendación que proporcionen búsquedas o filtros personalizados de acuerdo con la necesidad de búsqueda de los usuarios. En los buscadores (Google, Yahoo, etc.), la priorización y la personalización de la información se encuentra ausente, lo que ha incrementado la demanda de nuevos sistemas de recomendación que proporcionen búsquedas o filtros personalizados de acuerdo con la necesidad de búsqueda de los usuarios. Estos sistemas tienen la capacidad de predecir, si un usuario en particular preferiría un ítem o no; en función del perfil del usuario. En este proyecto se identificó y justificó el enfoque de recomendación a implementar para el sistema de recomendación de literatura académica, el cual tiene como estructura y base una taxonomía facetada de cáncer de mama, en específico para el caso de diagnóstico y tratamiento, la cual se diseñó con base del Consenso Mexicano del Cáncer de Mama en su versión más reciente.

**Palabras Clave:** Sistemas de Recomendación, Taxonomías, Web Scraping, Recuperación de Información, Cáncer Mamario.

## Capítulo 1

### 1.1 Introducción

Actualmente acceder y recuperar artículos de investigación es una tarea difícil, debido a la gran cantidad de publicaciones disponibles en los diferentes repositorios ya sean públicos o privados relacionados con las diferentes áreas científicas, temas y subtemas por especialidad. Para citar un ejemplo, sobre COVID-19, se han publicado más de 4,810,000 artículos desde el inicio de la pandemia y en particular podemos mencionar que, desde el mes de enero del presente año, se han publicado 65,200 artículos, cantidad que se duplica en promedio cada 20 días, resultando que sea imposible para los virólogos y otros especialistas seguir el ritmo de los nuevos artículos científicos sobre el tema. Los sistemas de recuperación de información, como Google, han resuelto parcialmente este problema, pero la priorización y la personalización (donde un sistema asigna el contenido disponible a los intereses y preferencias del usuario) de la información está ausente, lo que ha incrementado la demanda de los Sistemas de Recomendación (SR). Los SR son aplicaciones de la inteligencia artificial (IA) que están a la vanguardia en la investigación en recuperación de información y sistemas de filtrado de información. A partir de su aparición en 1979 (Rich, 1979), se han desarrollado para recomendar, en tiempo real, información relevante para los usuarios típicamente sin la necesidad de una consulta de búsqueda explícita (Jugovac, Jannach, & Lerche, 2017); En este proyecto consideramos un tipo particular de SR, denominados Sistemas de Recomendación de Literatura Académica (SRLA) este tipo de sistemas se caracteriza por ser un subtipo de SR dedicado a la recomendación de artículos académicos. Este dominio, tiene relativamente pocos usuarios en comparación con la gran cantidad de artículos de investigación en línea. Podemos citar dos desafíos en este tipo de sistemas: 1) La cantidad de dimensiones a considerar para llevar a cabo las recomendaciones: temas, subtemas, factor de impacto de la publicación, indexaciones, número de citas y fechas de publicación entre otras. 2) Acceso de los artículos, ya que pueden encontrarse en bancos de datos no públicos que requieren permisos de acceso, limitándose así el espectro de los artículos disponibles a recomendar para satisfacer los requerimientos de los usuarios.

En este proyecto de investigación nuestro objetivo es el de diseñar un SRLA en el ámbito de la medicina y en particular proponer un prototipo de SRLA para la recomendación de artículos de investigación en Cáncer de Mama en específico aquellos de diagnóstico y tratamiento.

## **1.2 Motivación y Objetivos**

Como se ha mencionado, el uso de los sistemas de recomendación se encuentra en incremento, sin importar la disciplina. En el caso de sistemas de recomendación enfocados en el cáncer de mama, se pueden acceder a pocos, sin embargo, abarcan todo el concepto en general, sin tener funcionalidades para acceder a temas específicos, esto genera el motivo principal para desarrollar este proyecto.

### **1.2.1 Objetivo General**

Desarrollar un prototipo de Sistema de Recomendación de Literatura Académica para artículos científicos médicos en el ámbito de cáncer de mama, específicamente diagnóstico y tratamiento, basado en taxonomías dinámicas.

### **1.2.2 Objetivos Específicos**

1. Definir el tipo de sistema de recomendación a realizar.
2. Analizar y establecer los criterios para el orden de la presentación de resultados a los usuarios.
3. Realizar una investigación sobre las metodologías más adecuadas para la recuperación de artículos académicos.
4. Analizar y definir la taxonomía de dominio en Tratamiento y Diagnóstico de cáncer de mama.
5. Realizar un prototipo de SRLA utilizando Taxonomías dinámicas.

## 1.3 Metodología

Durante los últimos 30 años las notaciones de modelado y las herramientas debieron ser la clave para garantizar de manera contundente que el desarrollo de software se realizara con éxito, sin embargo, esto no ha sucedido como se esperaba debido a que la metodología de desarrollo no se tomaba en cuenta. Se pueden tener las mejores notaciones y e infinidad de herramientas, sin embargo, sin directivas para su aplicación, no sirven de nada. En los últimos 15 años, el interés en las metodologías ha aumentado significativamente, por tanto, el proceso de desarrollo marca un gran énfasis en el control del proceso por medio de fuertes definiciones de roles, actividades, modelado y documentación lo más detallado posible, convirtiéndose en la estructura tradicional que toda metodología de desarrollo de software maneja en la actualidad (H. Canós, Letelier, & Penadés, 2007).

El objetivo de las metodologías ágiles se centra en permitir el trabajo rápido que los equipos de desarrollo de software desempeñan al realizar un proyecto, y adaptarse a los cambios que puedan surgir espontáneamente sin afectar las fechas de entrega. Esto claramente es una alternativa que se ofrece a comparación de las metodologías tradicionales, dado que estos son rígidos y dirigidos por la documentación obtenida de cada una de las actividades desarrolladas.

Por medio de una metodología de diseño iterativa e incremental, las fases a seguir se definen a continuación:

1. Investigación formativa en fuentes secundarias sobre sistemas de recomendación en general y los semánticos en particular para su uso en los sistemas de recomendación de artículos científicos.
2. A partir de los resultados de la investigación formativa proponer y justificar el enfoque de recomendación basado en contenido y taxonomías.
3. Documentar Taxonomía: Los procesos que se llevaran a cabo conforme a esta metodología son: a) Identificar el propósito y el ámbito de la taxonomía, b) Construir la taxonomía y d) Documentar la taxonomía. Para el diseño de la taxonomía nos basaremos en el Consenso Mexicano sobre Diagnóstico y Tratamiento del Cáncer Mamario (Cárdenas S., 2021).
4. Definir la arquitectura del sistema de recomendación de artículos científicos: a) Establecer los algoritmos del rastreo y búsqueda de PDF académicos, b) Algoritmos para realizar el proceso de búsqueda de nuevos artículos de forma periódica c) Creación del corpus.

## Capítulo 2

### 2.1 Estado del Arte

#### 2.1.1 Antecedentes del Proyecto

Los SR son sistemas de filtrado de información (Rich, 1979), que filtran dinámicamente, fragmentos de información, de acuerdo con las preferencias, intereses o comportamiento observado del usuario sobre el objeto (Jugovac, Jannach, & Lerche, 2017). Los SR, tiene la capacidad de predecir, si un usuario en particular preferiría un *ítem* o no; en función del perfil del usuario. En las bibliotecas científicas, los SR permiten ir más allá de las búsquedas en el catálogo. Por lo tanto, no se puede dejar de enfatizar la necesidad de utilizar técnicas de recomendación eficientes y precisas dentro de un sistema que proporcionará recomendaciones relevantes y confiables para los usuarios.

En la investigación académica, los Sistemas de Recomendación de Literatura Académica (SRLA) pueden ayudar a los investigadores, a encontrar rápidamente los artículos que necesitan (Trappey, Trappey, Wu, Fan, & Lin, 2013) Por ejemplo, a investigadores jóvenes con experiencia editorial limitada, los SR pueden recomendarles artículos recientes y artículos clásicos de áreas relacionadas para ampliar sus horizontes e intereses de investigación. Por el contrario, para los investigadores senior con registros de publicación más sólidos, los SR, le presentan principalmente artículos que se alinean con sus intereses de investigación (Sugiyama & Kan, 2010).

Las técnicas de recomendación pueden tipificarse en cuatro categorías según su método de filtrado: *basado en contenido* (CBF), *colaborativo* (CF), *basados en conocimiento* (KB) e *híbridos*. Cada método tiene su propia lógica subyacente (Xia, Liu, Lee, & Cao, 2016). CBF considera principalmente la preferencia histórica de los usuarios y su biblioteca personal para extraer y construir el modelo de interés de los usuarios, que se denomina perfil de usuario (Sun, Ma, Liu, & Miao, 2014). Luego, CBF extrae palabras clave de los artículos candidatos y calcula la similitud de las palabras clave extraídas de perfiles de usuario y candidatos documentos. Después de clasificar la similitud, se recomendarán a los usuarios artículos con alta similitud. CF se centra en la acciones o valoraciones sobre los elementos de otros usuarios cuyos perfiles son similares, vecinos cercanos (VC) del usuario (Pera & Ng, 2014), la hipótesis es que los usuarios que tienen un interés similar en el pasado, probablemente estarían de acuerdo también en el futuro.

Para el método híbrido, SR generalmente utilizar CBF y CF. Existen también, los sistemas de recomendación semánticos (SRS), dirigidos a dominios específicos y se basan en la interpretación semántica de las necesidades de información del usuario. La idea central de estos sistemas es utilizar el conocimiento ontológico para describir los elementos y tener una representación profunda y estructurada de su contenido. Los SRS forman parte de los SR-KB, con la diferencia de que los KB usan una base de datos de información generada por el mismo sistema o por los usuarios del sistema y los semánticos trabajan con información dotada de significado, haciendo uso de ontologías.

Por el lado de SR enfocados en publicaciones científicas, en el trabajo llamado “A content-based recommender system for computer science publications” (Wang, Liang, Xu, Feng, & Guan, 2018) ofrecen ayuda a los autores por medio de su trabajo para que puedan decidir donde enviar sus manuscritos, de tal manera que presentan revistas y conferencias basadas en contenido por medio de un sistema de recomendación en informática, así como su servicio web en <http://www.keaml.cn/prs/>. Este sistema recomienda revistas o conferencias adecuadas con un orden de prioridad basado en el resumen de un manuscrito. empleando un rastreador web para actualizar el conjunto de información y el modelo de aprendizaje, con el objetivo de lograr una respuesta interactiva en línea, los autores del artículo proponen un modelo híbrido basado en la selección de características chisquare y regresión softmax.

En el estudio “Designing Recommender System: Conceptual Framework and Practical Implementation” (Jallouli, Lajmi, & Amous, 2017), se presenta un marco de trabajo para desarrollar y evaluar un sistema de recomendación, describen de igual manera un modelo del proceso de preparación para el conjunto de datos, así como la explicación de varios algoritmos, todo esto con el objeto de poder desarrollar cuatro tipos de Sistemas de Recomendación, línea base, social, contextual y socio-contextual.

En el trabajo “Scientific Paper Recommendation: A Survey” (Bai, et al., 2019) describen como los SR han tomado una gran importancia a nivel mundial, enfocándose en primera instancia gracias a las aplicaciones de comercio electrónico y diferentes comunidades de investigación. Posteriormente se centran en explicar porque el recomendar artículos científicos, ya que en primera instancia se tiene el tiempo ya que no a todos los usuarios les va a interesar un artículo clásico, algunos preferirán de reciente publicación. Este campo es bastante atractivo ya que el número de artículos académicos aumenta de manera exponencial.

En este trabajo también presentan los métodos de evaluación de diferentes SR, terminando con la mención de los problemas que los SRLA presentan, como son la escasez, la escalabilidad, la privacidad, el estándar de datos académicos unificados, entre otros.

En el estudio “A Study of Recent Recommender System Techniques” (Bansal & Baliyan, 2019), se estudian los diferentes enfoques de recomendación, sus limitaciones y su extensión, pues cuando se aplican estos enfoques pueden erradicar por completo el problema de la sobrecarga de información en la web. De igual manera realizaron una investigación sobre la base de datos Google Scholar para delinear la distribución temporal de las diferentes técnicas de recomendación que utilizan.

En “Métricas de similaridad y evaluación para sistemas de recomendación de filtrado colaborativo” (Mendoza Olguín, Laureano de Jesús, & Pérez de Celis Herrero, 2019) presentan un análisis comparativo entre algunas métricas de similitud y evaluación propuestas para los sistemas de recomendación basados en filtrado colaborativo; realizando pruebas sobre datasets comúnmente utilizados para determinar su eficiencia en producción de acuerdo con algunos sitios web de e-commerce.

En el trabajo “Estado del arte en los sistemas de recomendación” (González Escamilla & Marcellin Jacques, 2017) se definen a los SR y a cada uno de los tipos de SR de manera formal, explicando a detalle cómo es que se lleva a cabo el filtrado por cada tipo de SR de manera matemática, de igual manera realizan una comparación entre las distintas técnicas de filtrado que mencionan en su escrito.

De acuerdo con “Introduction to Recommender Systems Handbook” (Rici, Rokach, & Shapira, 2010), los Sistemas de Recomendación (SR) son herramientas y técnicas de software que proporcionan sugerencias de elementos que pueden ser de interés y de utilidad para un usuario. Estas sugerencias se encuentran relacionadas con varios procesos de toma de decisiones como que música escuchar, que noticias leer o que cosas comprar por medio de internet.

El término general utilizado para denotar lo que un SR recomienda a los usuarios se conoce como *item*. Comúnmente un SR se centra en un tipo específico de *item*, que coincide con el diseño e interfaz gráfica que se presenta al usuario.



El *core*, mejor conocido como el núcleo de la técnica de recomendación, es diseñado de acuerdo con las necesidades de los usuarios, para generar recomendaciones personalizadas que sean útiles y eficaces para cada tipo de *item*.

Una de las características principales de los SR es que están dirigidos hacia personas que carecen de la suficiente experiencia o preparación para evaluar la gran cantidad de *ítems* que un sitio web puede ofrecer, por ejemplo, un SR de recetas puede ayudar a usuarios a obtener una receta que se acople a sus necesidades o gustos.

Un usuario puede recibir diversas sugerencias de una búsqueda dado que las recomendaciones suelen ser personalizadas, sin embargo, también hay recomendaciones no personalizadas.

Las recomendaciones no personalizadas son más sencillas de generar, comúnmente suelen aparecer en revistas o periódicos, un ejemplo, serían las 10 mejores canciones del momento o los mejores libros publicados por una editorial. Aunque estas sugerencias pueden ser útiles en determinadas circunstancias, no suelen ser objeto de investigación de los SR.

De manera simple, las recomendaciones personalizadas suelen ofrecerse como listas clasificadas de ítems, al realizar esta clasificación, los SR tratan de predecir qué productos o servicios son los más adecuados, basándose en las preferencias y limitaciones del usuario, para que esto se logre, los SR tienen integrados diversos algoritmos para obtener las preferencias de los usuarios o para analizar la navegación o el historial de búsqueda.

En la tesis “Sistemas de recomendación semánticos para la compartición de conocimiento y la explotación de tesauros. Un enfoque práctico en el ámbito de los sistemas nutricionales” (Martín, 2016) se describen definiciones de lo que son los SR, los tipos de SR, pasando posteriormente a una explicación acerca de la Web semántica, tesauros, ontologías, mencionando definiciones, clasificación de ontologías, componentes puntuales de las ontologías y tesauros. Posteriormente se centran en los SR en entornos nutricionales, donde se toma en cuenta principalmente el modelado de los usuarios, la monitorización y el aprendizaje de acuerdo con sus preferencias alimenticias, presentando finalmente ejemplos de SR en el dominio de la nutrición. Finalmente se centran en mencionar la propuesta de los autores.



De acuerdo con el estudio “Research-paper recommender systems: a literature survey” (Beel, Gipp, Langer, & Breitinger, 2016), entre el año 2000 y 2015 se publicaron más de 200 artículos de investigación orientados en los Sistemas de Recomendación de Artículos de Investigación, estos artículos fueron sometidos a revisión y presentan estadísticas descriptivas de ellos, incluyendo definiciones generales, conceptos y enfoques generales. En este trabajo también hablan acerca de que un 55% de los enfoques de recomendación aplicaban un filtrado basado en contenido, mientras que un 18% aplicaba filtrado colaborativo, mientras que 16% de los trabajos revisados implementaban recomendaciones basadas en grafos, mientras que los restantes se sitúan en esquemas híbridos. Una vez hecho este conteo se centran los autores en realizar estadística y comparación de cómo es que ciertos enfoques implementados pudieron haber presentado ambigüedad en los resultados o incluso comprobaron si existían variaciones en el rendimiento de los enfoques aplicados. Los autores concluyen con que el 73% de los autores no publicaron más de un artículo de investigación al respecto, lo cual no ayuda a mejorar el panorama ya que se queda todo en una sola implementación y se da por terminada la investigación o difícilmente habilitan mejoras a sus SR.

En “Sistemas de Recomendación” (Pérez Almaraz & Close G., 2015) explican de manera general de lo que son los sistemas de recomendación, su importancia en distintas áreas de investigación, aplicaciones en el medio del entretenimiento, contenido, e-commerce o servicios. De igual manera escriben sobre la historia de los SR, las técnicas de SR: Filtrado Colaborativo, Basado en Contenido, Demográfico, Basado en Conocimiento, Basado en la Comunidad e Híbrido, concluyendo con ejemplos actuales, mencionando a Amazon, Netflix, etc.

En el artículo “The Architecture and Datasets of Docear's Research Paper Recommender System” (Beel, Langer, Gipp, & Nürnberger, 2014), presentan la arquitectura del sistema de recomendación que desarrollaron para artículos de investigación de su software de gestión de referencia Docear, acompañado de cuatro conjuntos de datos. En la parte de la arquitectura describen como rastrean los archivos en formato PDF, de igual manera generan modelos de usuario y los cálculos de recomendaciones basados en contenido. Esto con el fin de apoyar a investigadores y desarrolladores para que construyan sus propios Sistemas de Recomendación de artículos de investigación. Referente a los cuatro conjuntos de datos, estos contienen 9.4 millones de artículos académicos.

En el trabajo “Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation.” (Manouselis & Verbert, 2013) se estudian los enfoques aplicados por plataformas como Mendeley, CiteSeer y Google Scholar quienes presentan distintos enfoques para generar recomendaciones de artículos científicos, pero en particular se centran en aplicar un marco de evaluación por capas para adaptar la interacción de los componentes de un SR. Se trabajó sobre los datos de la plataforma Mendeley bajo el prisma de un marco de evaluación de capas. El principal beneficio de este enfoque de evaluación identifica primeramente la forma en que puede revelar si la elección de un valor de propiedad de datos puede afectar el rendimiento de los algoritmos, también ayuda a justificar si el modelado de los usuarios de criterios múltiples puede mejorar el rendimiento de los SR y de que maneras distintas.

En el estudio “A literature review and classification of recommender systems research” (Park, Kim, Choi, & Kim, 2012) mencionan una definición formal de los SR, pasando a explicar posteriormente una metodología de búsqueda, donde obtuvieron 164 artículos en los sistemas de recomendación de 31 revistas, clasificaron la literatura en ocho categorías por medio de técnicas de minería de datos, usando: Reglas de asociación, agrupación en clústeres, árbol de decisión, vecino más cercano k, análisis de enlace, red neuronal, regresión y otros métodos heurísticos. Sin embargo, inician clasificando de acuerdo con sus propios criterios, pasando primero por 2 evaluadores en la primera fase, para la segunda se verifica con otros dos verificadores del equipo. De los 164 artículos clasificados de 31 revistas se prosiguió a distribuir los artículos por año, del existiendo menos artículos del 200 al 2004, mientras que hubo un gran incremento del 2007 al 2009.

En la tesis “Sistemas de recomendación para webs de información sobre la salud” (Seguido Font, 2009) se relata la historia de los SR, continuando con definiciones, características y clasificaciones, por capítulo se mencionan cada uno de los tipos de SR actuales, sin embargo referente al tema de la salud, mencionan a la CYTED(Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo) que es quien se encarga de administrar los sitios para el área de la salud, describiendo su historia, a que se dedica y hace comparaciones sobre la calidad de los sitios, finalmente mencionan la importancia de los SR pero a manera de crítica de los sitios que se tienen en el CYTED.

Por otro lado, en “Sistemas de Recomendación Semánticos. Un análisis del estado de la cuestión” (Peis, Morales del Castillo, & Delgado López, 2008) hablan acerca de los sistemas de recomendación semánticos, en específico de aquellos que basan su funcionamiento sobre una base de conocimiento definida a través de un esquema de conceptos (como una taxonomía o un tesaurus) o una ontología, y que utilizan tecnologías de Web Semántica.

En el estudio “A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques” (Nilashi, Ibrahim, & Bagherifard, 2018) se describe el desarrollo de un nuevo método de recomendación híbrido basado en enfoques de filtrado colaborativo (CF), con el objetivo de resolver dos inconvenientes comunes de los Sistemas de Recomendación, la escasez y la escalabilidad, por medio de técnicas de reducción de dimensionalidad y ontología. La ontología se usa para mejorar la precisión de las recomendaciones en la parte CF. En el CF parte, también utilizan una técnica de reducción de dimensionalidad llamada Descomposición de Valor Singular (SVD), para encontrar elementos y usuarios más similares en cada grupo de elementos y usuarios que pueden mejorar significativamente la escala capacidad del método de recomendación. Evaluando el método en dos conjuntos de datos del mundo real para mostrar su efectividad y comparar los resultados con los resultados de los métodos en la literatura.

En “Filtrado Colaborativo y Sistemas de Recomendación” (Galán Nieto, 2007), se habla acerca de los sistemas de recomendación basados en algoritmos de filtrado colaborativo, utilizando valoraciones de usuarios sobre elementos de un conjunto total para predecir valoraciones en los elementos restantes y recomendar aquellos que tengan la mayor valoración predicha. Describe a los sistemas de filtrado colaborativo, de igual manera presenta información acerca del uso de la librería de filtrado “Taste” que contiene implementaciones de varios algoritmos de filtrado colaborativo.

## 2.1.2 Sistemas de Recomendación

### 2.1.2.1 Definición Formal

En “Recommending and Evaluating Choices in a Virtual Community of Use “ (Hill, Stead, Rosenstein, & Furnas, 1995) definen formalmente a los sistemas de recomendación como:

**U:** Conjunto de Usuarios.

**I:** Conjunto de posibles ítems (o elementos) a recomendar.

**u:** Usuario sobre el cual se realiza la recomendación.

**i:** ítem para el que se quiere generar la predicción de la preferencia de u.

**fu:** Función de utilidad que mide la utilidad de i para u. Habitualmente es la ratio.

Por medio de las siguientes ecuaciones se puede formular un SR:

$$fu : U \times I \rightarrow R$$
$$\forall u \in U, \exists i' \in I \mid i'_u = \max_{i \in I}(fu(u, i))$$

La función de utilidad suele ser la puntuación dada por el usuario al ítem, la cual comúnmente es utilizada como información de retroalimentación en los SR.

### 2.1.2.2 Algoritmos de filtrado

- Algoritmos de filtrado colaborativo basados en memoria, o algoritmos de vecinos cercanos (*Nearest Neighbour*), emplean técnicas estadísticas para encontrar a vecinos, es decir usuarios con un historial de valoraciones sobre los elementos similar al usuario actual.
- Algoritmos de filtrado colaborativo basados en Modelo, desarrollan en primera instancia un modelo de los ratings por usuario, tratan el problema como un problema de predicción estadística y calculan el valor esperado para cada ítem en función de los ratings anteriores, utilizan distintos algoritmos de aprendizaje *Clustering* o redes neuronales como las Redes de Funciones de Base Radial (RBFN).

### 2.1.2.3 Técnicas de Recomendación

A continuación, se describirán los diferentes tipos de SR de acuerdo con una taxonomía proporcionada por el trabajo “Collaborative recommendation: A robustness analysis” (O'Mahony, Hurley, Kushmerick, & Silvestre, 2004), siendo una forma clásica para distinguir a los sistemas de recomendación de acuerdo con distintos enfoques.

- **Basados en contenido (Content-Based):** Este tipo de SR aprende a recomendar *ítems* que son similares a los que le gustaron a un usuario en el pasado. La similitud de los *ítems* se calcula a partir de las características asociadas a los artículos comparados. Por ejemplo, si un usuario valoró positivamente una película, el sistema conoce a que genero pertenece y aprende a recomendar otras películas que compartan el mismo género.
- **Filtrado colaborativo (Collaborative filtering):** Es la implementación más sencilla en los SR, este enfoque realiza recomendaciones al usuario activo de todos los *ítems* que a otros usuarios con gustos similares les haya gustado en el pasado. En este caso la similitud de gustos entre dos usuarios se calcula en función de la similitud en el historial de valoraciones de los usuarios. Es la técnica más popular e implementada en los SR.
- **Demográfico:** Este SR recomienda *ítems* en función del perfil demográfico del usuario, existen sitios web que generan recomendaciones de acuerdo con el idioma o al país de donde se encuentre el usuario.
- **Basados en conocimiento (Knowledge-based):** Recomiendan *ítems* basados en el conocimiento específico del dominio sobre como ciertas características de un *ítem* satisfacen las necesidades y preferencias de los usuarios, donde en última instancia se toma en cuenta como el *ítem* es útil para el usuario. Existe una variante de este tipo de SR, los llamados basados en casos (case-based), los cuales su función de similitud estima en qué medida las necesidades del usuario coinciden con las recomendaciones, se modela una descripción del problema y se regresa un *ítem* como solución.

- **Basado en la comunidad (Community-based):** Este tipo de SR recomienda ítems de acuerdo con las preferencias de los amigos del usuario activo. Los datos que se han obtenido sugieren que las personas tienden a confiar más en las recomendaciones de sus amigos que en las de personas con gustos similares anónimas. Este SR modela y adquiere información sobre las relaciones sociales de los usuarios y las preferencias de sus amigos, teniendo una mayor relevancia en las redes sociales.
- **Sistemas de Recomendación Híbridos (Hybrid recommender systems):** Estos SR se basan en la combinación de las técnicas mencionadas anteriormente. Principalmente combina técnicas de A y B, donde intenta utilizar las ventajas de A para solucionar los inconvenientes de B. Por ejemplo, los métodos de Filtrado Colaborativo tienen problemas si se agregan nuevos *ítems* ya que no tienen alguna valoración, por otro lado, esto no limita los métodos de los SR Basado en contenido, ya que la predicción de *ítems* nuevos se basa en sus características que suelen estar disponibles. Entonces fusionando los métodos de estos dos tipos de SR se crea uno nuevo con mayores capacidades y nuevas características.
- **Sistemas de Recomendación Semánticos (SRS):** Al existir brechas y problemas sin resolver en los SR anteriormente mencionados se plantea una nueva estrategia para abordar y dar solución a estos problemas, que en los últimos años ha tenido mayor relevancia, la solución comúnmente se presenta fusionando métodos de la web semántica con los SR. Las ontologías aportan el conocimiento sobre el dominio clasificando los *ítems* en categorías, a partir de ello se construyen relaciones de similitud entre los objetos en cuestión.

Ahora bien, de acuerdo con la tesis “Sistemas de recomendación de contenidos para libros inteligentes” (Núñez V., 2012) los SRS se pueden clasificar en tres tipos:

- **Sistemas basados en ontologías o esquemas de conceptos:** Estos SRS realizan sus recomendaciones por medio de ontologías con el fin de representar la información, modelar los perfiles de los usuarios y sus elementos. Utilizan tecnologías y métodos de la web semántica para así mostrar el contenido en la web y poder consultarlo por medio de RDF (Marco de Descripción de Recursos) y SPARQL (Lenguaje de consulta para RDF).

- **Sistemas adaptables al contexto:** Toman en cuenta una serie de factores que determinen la situación o el contexto en el que se encuentra el usuario, con el fin de que las recomendaciones se adapten a esa situación o contexto. Los factores por tomar en cuenta son el lugar donde se encuentra, el tiempo, nivel de experiencia e incluso el dispositivo que está usando el usuario al momento de recibir la recomendación.
- **Sistemas basados en redes de confianza:** Crean redes de confianza entre los diferentes componentes del sistema. De igual manera agregan filtros de datos adicionales a los sistemas adaptables al contexto, con el fin de garantizar fiabilidad y una alta precisión en las recomendaciones mediante la creación de redes de confianza entre los componentes del SR.

## Capítulo 3

### 3.1 Análisis y especificación de requisitos

Teniendo definidas tres preguntas de investigación, a partir de responderlas se definió el rumbo de este proyecto de tesis. Para darles respuesta se realizó una revisión de la literatura para comprender y definir conceptos básicos y teóricos referentes a los Sistemas de Recomendación (SR), taxonomías y recuperación de información, a continuación, se presentan las preguntas con sus respectivas respuestas:

1. ¿La filosofía de los SRS con enfoque de recomendación basado en taxonomías es el más adecuado para la implementación de Sistemas de Recomendación de Literatura Académica?

En efecto, es el más adecuado para la implementación de SRLA, ya que por medio de la taxonomía se puede clasificar el conjunto de datos (Artículos Académicos) y a su vez, esta clasificación ayudaría a obtener una mejor respuesta y más acertada a las necesidades de búsqueda de los usuarios. Además de que dada la flexibilidad de la taxonomía y a que sus objetos se encuentran multclasificados se pueden afinar búsquedas y el usuario podría navegar por las características y obtener un mejor resultado.

A pesar de revisar diversa documentación en específico se utilizó como parte fundamental de la investigación el libro "Introduction To Recommender Systems Hanbook" (Rici, Rokach, & Shapira, 2010), reforzando conocimientos sobre ¿Qué son los SR?, Funcionalidad, Conjunto de datos a manejar, Técnicas de recomendación, Técnicas de Aplicación y evaluación. De igual manera como en lo SRS se utilizan las tecnologías de la web semántica, se buscaron taxonomías u ontologías referentes al cáncer de mama, para usarlos como base para la creación de la taxonomía que utilizaría el SRS para clasificar los diferentes artículos a trabajar.

2. ¿Cuáles son las métricas de evaluación adecuadas para medir la calidad de las recomendaciones?

- Se toma en cuenta la cantidad de citas realizadas hacia un artículo, para determinar la relevancia de este.
- Además de tomar en cuenta el factor de impacto y el cuartil al que pertenece un ítem.



3. ¿Qué metodología es la adecuada para para la extracción de registros de las bases de datos bibliográficas y como garantizar la construcción del depósito incremental de artículos de investigación?

De acuerdo con la investigación y pruebas realizadas, se determinó utilizar herramientas referentes al Web Scraping (Python - Selenium, SerpApi) utilizando en general un API para la obtención de información el cuál funciona bajo técnicas de *scrap*, así como una base de datos relacional para representar y manipular la estructura de la taxonomía, teniendo como resultado el guardado de la información relevante y el rastreo de los artículos nuevos a indizar.

## 3.2 Casos de Uso

### 3.2.1 Actores

Usuario General

Para esta versión se cuenta con un único usuario que interactuará con el sistema, el cual deberá iniciar sesión en el sistema para poder realizar búsquedas, de igual manera podrá visualizar el historial de búsquedas que ha realizado, así como podrá ver información relevante del sitio, como la cantidad de documentos indexados acerca del Cáncer de mama, el total de usuarios registrados en el sitio e incluso el número total de búsquedas que realizó el usuario desde que ingreso al sitio. El inicio de sesión se toma como base para poder saber que usuario búsquedas y cuales fueron, con el fin de poder devolverle información relevante acuerdo a sus búsquedas recientes, si el usuario no está registrado en el sistema podrá hacerlo o incluso podrá solicitar que se restablezca su contraseña.

En la Figura 1, se puede observar el diagrama de caso de uso para este usuario, teniendo que el usuario podrá Registrarse en el sistema de tal manera que esto le permitirá *Registrar Datos Perfil*, para identificar al usuario. Una vez realizado esto, el usuario podrá *Realizar Búsquedas* y así *Acceder a la Información*. Por otro lado, si el usuario ya está registrado entonces podrá Iniciar Sesión para así *Realizar Búsquedas* y *Acceder a la Información*. En caso de que el usuario no recuerde los datos para iniciar sesión, podrá *Recuperar Contraseña* de tal manera que le permita el acceso al sitio.

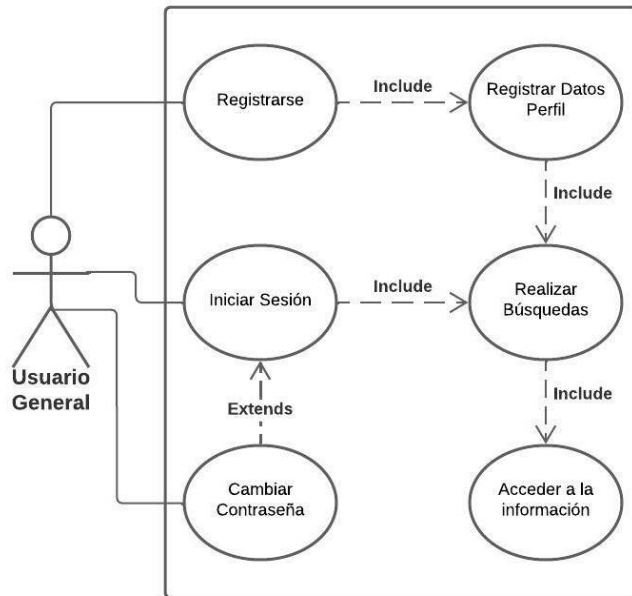


Figura 1 - Caso de Uso Usuario General

## Capítulo 4

### 4.1 Diseño de la Taxonomía

Las taxonomías permiten tener objetos multclasificados lo que hace que la recuperación de la información sea más precisa, pues ayudan al usuario a buscar a través de las etiquetas de las clases y subclases pertenecientes de la estructura jerárquica.

Se realizó una búsqueda exhaustiva con el fin de encontrar taxonomías u otras estructuras que clasificarán el Cáncer de mama, entre los recursos más importantes que se encontraron se puede apreciar en la Figura 2, una clasificación médica de los tipos de cáncer de mama, sin embargo, solamente identifica el tipo de cáncer, sin proveer más información de los términos relacionados a cada uno de ellos.

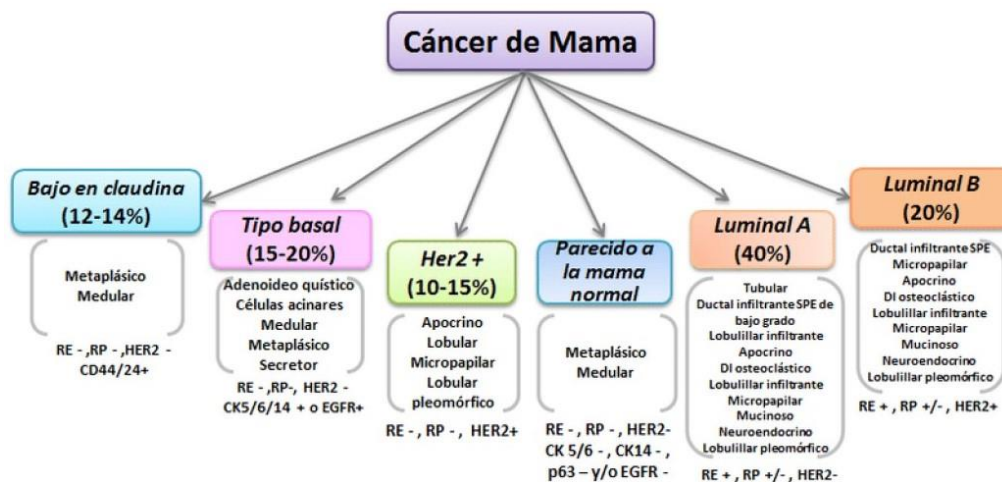


Figura 2 - Clasificación médica cáncer de mama

Por otro lado, en la Figura 3, se tiene una estructura un tanto más específica, teniendo una clasificación del cáncer de mama, pero para los objetivos del proyecto se encuentra muy limitada en cuanto a términos ya que se requiere obtener información más específica.

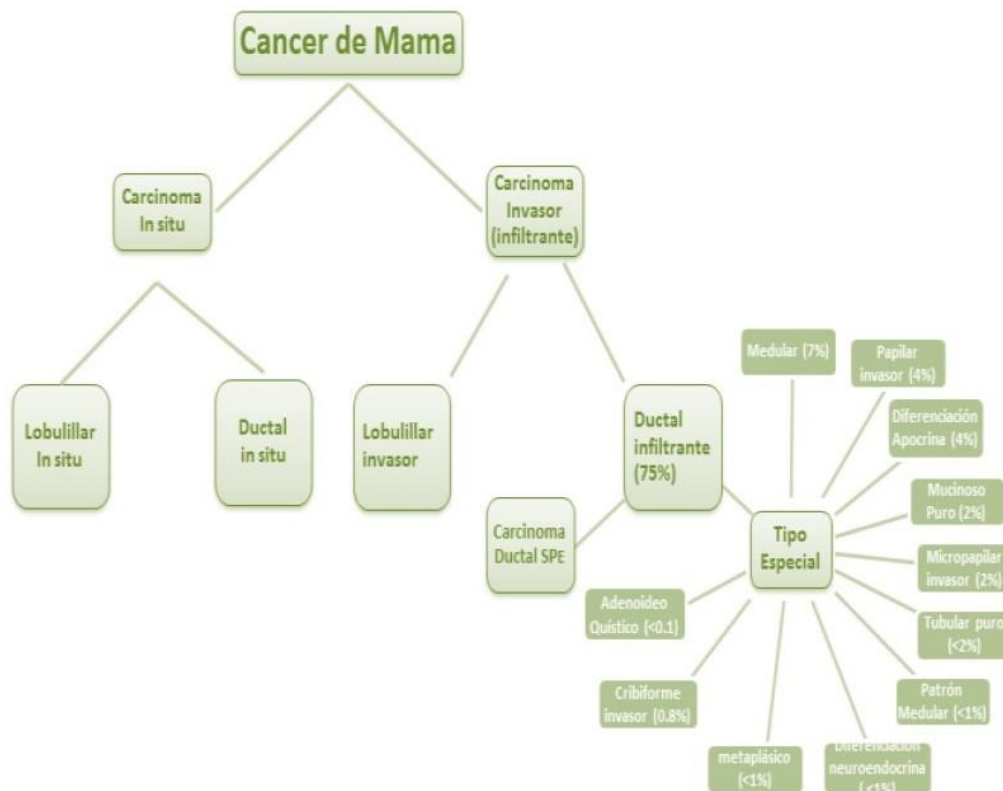


Figura 3 - Clasificación del Cáncer de Mama

Al no existir como tal una taxonomía digital que se pueda utilizar, se planteó utilizar el formato del Consenso Mexicano sobre Diagnóstico y Tratamiento del Cáncer Mamario en su versión más reciente (Cárdenas S., 2021), para construir la taxonomía ideal para este proyecto, ver Figura 4.

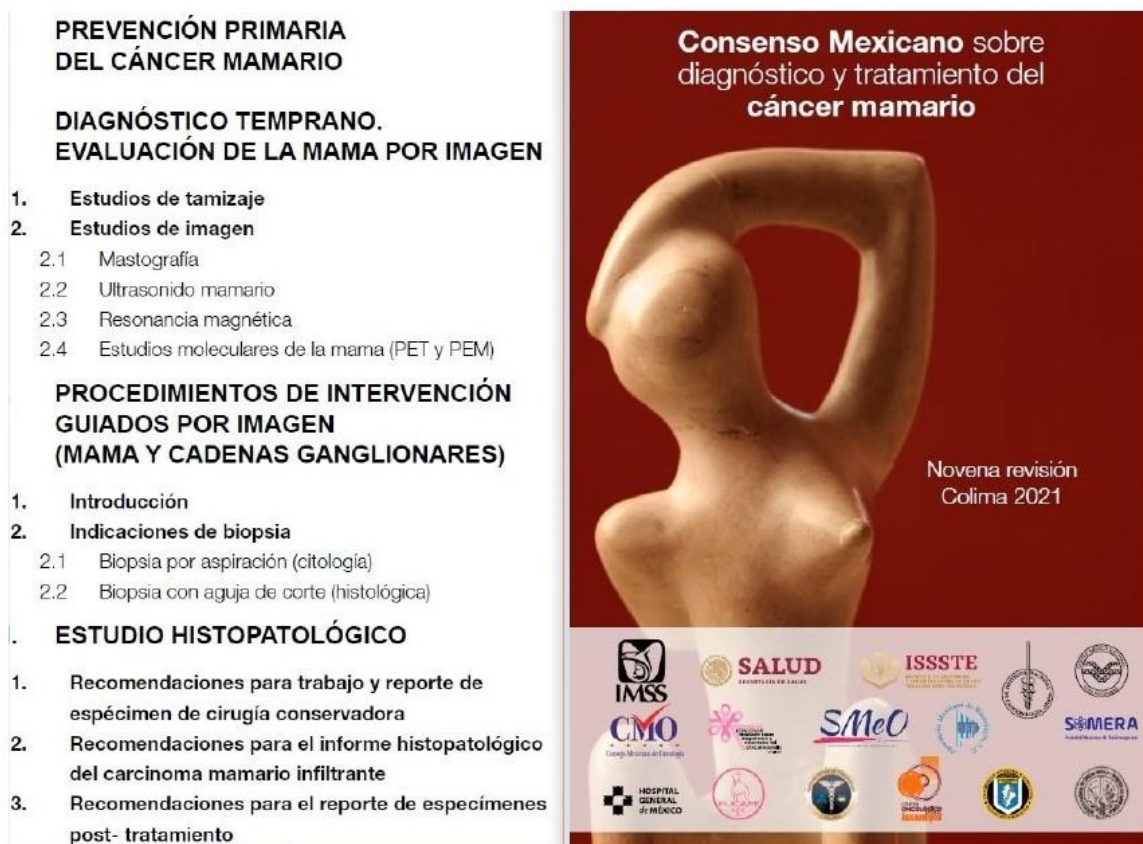


Figura 4 - Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario.

Fue en 1994, en la ciudad de Colima que se presentó el primer Consenso Nacional sobre Diagnóstico y Tratamiento del Cáncer Mamario, siendo desde entonces una de las guías más importantes a nivel nacional para oncólogos y otros médicos con especialidades relacionadas. Para que los temas tratados estén actualizados, se han realizado nueve sesiones para validar la información y de ser necesario agregar o modificar temas de acuerdo con los descubrimientos o avances en la medicina relacionada. Esto ha sido publicado en diversas revistas especializadas y gracias a esto, han tenido gran difusión, prácticamente la mayoría de los oncólogos del país están informados de las conclusiones del Consenso y le utilizan como herramienta de apoyo para tomar de decisiones en la práctica.

La revisión del 2021 se realizó de manera virtual el 20 y 30 de Enero, revisando los avances más recientes en el campo de la prevención, diagnóstico y tratamiento del cáncer mamario, siendo la herramienta más actualizada del tema. Se convocaron a 105 médicos reconocidos a nivel nacional de todas las instituciones y especialidades relacionadas con esta enfermedad, trabajando en conjunto por equipos se analizó la información de cada área con el fin de presentarla posteriormente y se aprobara para la publicación. Para esta edición añadieron un tema importante “Covid-19 y cáncer de mama” debido a la pandemia y por el impacto que ha llegado a tener en los pacientes.

Se espera que esta versión contribuya como guía a la comunidad médica y a los oncólogos con el fin de ofrecer a los pacientes con esta enfermedad, un diagnóstico preciso y un tratamiento óptimo y actualizado. Por eso mismo, dada la constante actualización de la información y su impacto en el área, se tomó la decisión de tomar al consenso como base para el diseño e implementación de nuestra taxonomía.

A partir de la forma en la que el consenso presenta su contenido fue como se definió la estructura de la taxonomía, del contenido se extrajeron los términos relevantes a utilizar y a partir de la estructura en conjunto con los términos obtenidos, se definieron las facetas y sus términos hijos.

Como se puede apreciar en la Figura 5, se tiene una representación de una Taxonomía T, tiene como facetas Consenso, Tipo Documento, Locación documentos, donde estos tienen relación con varios conceptos y se pueden anidar, mientras que a cada uno de ellos se le indizan documentos tomando en cuenta que el contenido del documento sea relevante para el concepto.

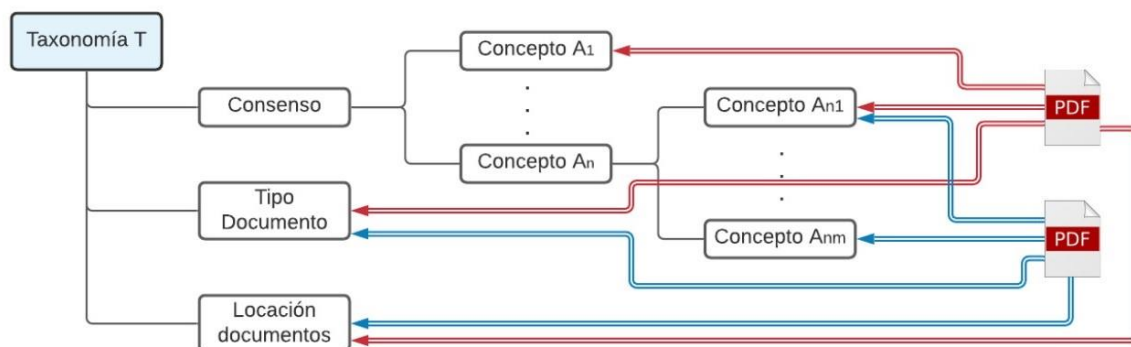


Figura 5 - Estructura general

La taxonomía de diagnóstico y tratamiento del cáncer de mamá servirá como base para definir y estructurar conceptos relacionados al diagnóstico y tratamiento en específico, para que en el SRLA dada una búsqueda sean devueltos como resultado aquellos documentos que realmente hablen de un concepto en cuestión de la búsqueda realizada.

#### 4.1.1 Proceso de creación

En la figura 6, se puede observar el diagrama de flujo que se tomó como base para la elaboración de la taxonomía, partiendo de que se define el concepto padre, este se inserta y se verifica si este concepto va a ser un concepto hoja, en caso afirmativo, se termina de trabajar con esa rama y se continúa con los demás conceptos, pero en caso de que no sea un concepto hoja se asignan los conceptos que serán sus hijos modificando la jerarquía que se tenía y para cada concepto nuevo se valora si serán o no conceptos hoja, de esta manera, una vez definiendo los conceptos padres, que serán las facetas de la taxonomía, se le relacionan los conceptos hijo.

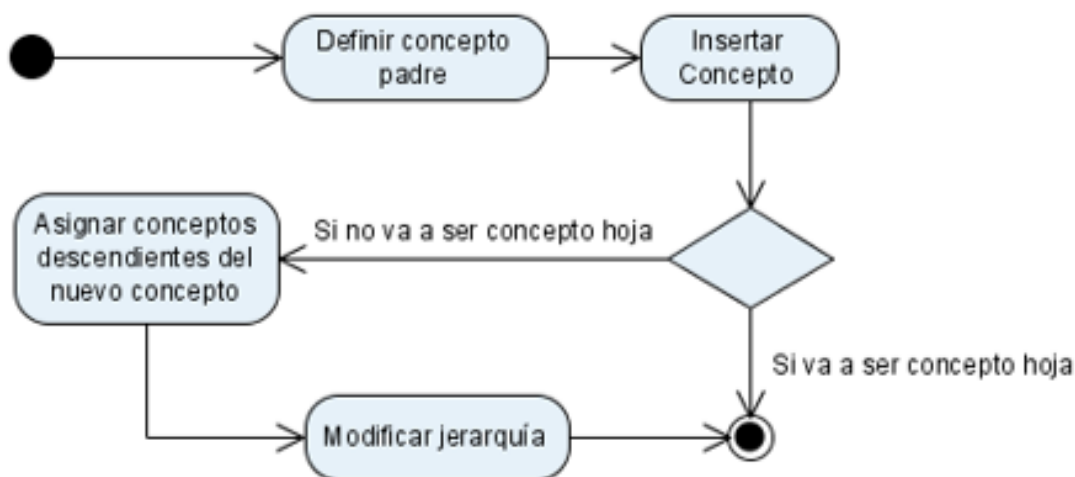


Figura 6 - Diagrama de flujo para la creación de la Taxonomía



### 4.1.2 Conceptos

A continuación, en la Tabla 1, se pueden observar los conceptos pertenecientes a la Taxonomía que se elaboró, teniendo del lado izquierdo el nombre del concepto y del lado derecho su descripción.

**Tabla 1. Diccionario de Conceptos de la Taxonomía**

Concepto	Descripción
Hormonoterapia Neoadyuvante	Recomendada en mujeres postmenopáusicas con receptores hormonales positivos y etapas II-III o en pacientes en quienes la toxicidad de la quimioterapia no sea aceptable o que padecen múltiples comorbilidades.
Quimioterapia Neoadyuvante	Puede reducir el tamaño de grandes tumores, y permitir a las pacientes que responden favorablemente ser buenas candidatas para la conservación de la mama, y así realizar mastectomía parcial, seguida de radioterapia.
Rehabilitación Física	Los avances en los tratamientos y el incremento de la supervivencia de los pacientes con cáncer de mama demandan que los métodos de rehabilitación sean cada vez más efectivos para lograr una mejor calidad de vida, tanto en los supervivientes de la enfermedad, como en pacientes en etapa terminal.
Terapias Blanco Neoadyuvantes	El tratamiento neoadyuvante recomendado se basa en 6-8 ciclos de quimioterapia ya que se asocian con mayores posibilidades de RPC.
Manejo Temprano	Los conocimientos actuales sobre las causas del cáncer de mama son insuficientes, por lo que la detección temprana sigue siendo el punto más importante de la lucha contra esta enfermedad.
Manejo Quirúrgico Primario	Está indicado para aquellas pacientes con cáncer de mama temprano. Este puede ser con cirugía conservadora o mastectomía total, independientemente del manejo quirúrgico de la axila.
Mastectomía	Término médico para la extirpación quirúrgica de una o ambas mamas de manera parcial o completa.



Mastectomía preservadora de complejo areola pezón	Esta técnica preserva todo el recubrimiento natural del complejo areola-pezón, y está orientada a crear un resultado estético más cercano al estado natural que las técnicas de reconstrucción de la mama.
Mastectomía Preservadora de piel	Conserva la máxima cantidad de piel y el surco submamario; facilita la reconstrucción inmediata y se consigue mejor simetría, aspecto natural de la mama y mayor satisfacción de la paciente.
Mastectomía radical	La mastectomía radical es un procedimiento quirúrgico que implica la extirpación de la mama, el músculo pectoral subyacente y los ganglios linfáticos de la axila como tratamiento para el cáncer de mama.
Mastectomía radical modificada	Una mastectomía radical modificada combina una mastectomía simple con la extirpación de los ganglios linfáticos axilares (disección de los ganglios linfáticos axilares).
Mastectomía simple	En este procedimiento, el cirujano extirpa todo el seno, incluyendo el pezón, la areola y la piel.
Mastectomía reductora de riesgo MRR	Opción de intervención en mujeres con alto riesgo para desarrollar cáncer de mama.
Mastectomía Reductora de Riesgo Contralateral MRRC	Se define como la mastectomía del lado sano, en una mujer con cáncer de mama unilateral.
Cirugía Conservadora	Escisión completa del tumor primario con margen patológico negativo.
Cirugía oncoplástica	La cirugía oncoplástica es un conjunto de técnicas quirúrgicas, que permite efectuar resecciones proporcionalmente mayores, con un resultado estético satisfactorio.
Abordaje reconstructivo para defectos parciales	Reparación de defectos de mastectomía parcial es más adecuado para pacientes con mamas grandes.
Reconstrucción de defectos parciales	La remodelación del tejido, antes de la radioterapia permite el uso de técnicas con tejido local (oncoplástica).
Técnicas reconstructivas para defectos parciales	Las técnicas de reconstrucción a menudo están influenciadas por el momento de la reconstrucción y la asociación a radioterapia.

Reconstrucción directa	Este proceso es más recomendado en escenarios de mastectomía reductora de riesgo, y para aquellas con cáncer de mama en estadio temprano.
Reconstrucción con expansor o implante	Este tipo de reconstrucción implica la colocación de un expansor retro muscular en el primer tiempo quirúrgico, infiltraciones para expansión en consultorio y al lograr el volumen deseado, la eventual colocación de un implante definitivo.
Reconstrucción con expansor + matriz dérmica	Se utiliza una matriz dérmica acelular para cubrir el polo inferior del expansor y el músculo pectoral mayor cubrirá el polo superior.
Reconstrucción con colgajos	El músculo latissimus dorsi y el colgajo perforante de la arteria toracodorsal, desempeñan funciones importantes en la reconstrucción post mastectomía. Estos colgajos son buenas opciones para las pacientes obesas, para los cuales no siempre es seguro realizar una reconstrucción con implantes, especialmente cuando no se requieren grandes volúmenes.
Colgajos pediculados	Estos colgajos son buenas opciones para las pacientes obesas, para los cuales no siempre es seguro realizar una reconstrucción con implantes, especialmente cuando no se requieren grandes volúmenes, también son adecuados para pacientes que se han sometido a radioterapia y que tienen defectos parciales.
Colgajo libre abdominal	En la práctica clínica se realiza la reconstrucción mamaria con colgajo libre del tejido abdominal bajo, mejor conocido como colgajo "DIEP" por sus siglas en inglés (Deep Inferior Epigastric Perforator).
Colgajo libre no abdominal	Algunas opciones para esta alternativa incluyen al colgajo libre de gracilis TUG (oblicuo, transverso, vertical), colgajo de perforante de arteria glútea superior o inferior (SGAP, IGAP), colgajo libre de perforante de arteria profunda femoral (PAP), colgajo de perforante de arteria lumbar (LAP), entre otros.
Transferencia o injerto graso	La transferencia de grasa es un método cada vez más popular para perfeccionar las reconstrucciones mamarias, es un complemento útil y seguro para la reconstrucción mamaria.

Tratamiento quirúrgico de la axila	En cáncer invasor, la evaluación axilar es parte fundamental de su manejo; el objetivo primario es la información pronóstica que brinda el estado ganglionar.
Dissección axilar	Es la forma clásica de tratar la axila en las pacientes con cáncer de mama, la idea de la dissección axilar es tanto lograr el control local de la axila como de saber si la axila está tomada por el tumor.
Tratamiento Hormonal	El tratamiento hormonal puede alcanzar a las células cancerosas en casi cualquier parte del cuerpo y no sólo en el seno. Se recomienda para las mujeres con tumores que son receptores de hormonas positivos. Sin embargo, no ayuda a las mujeres cuyos tumores no tienen receptores hormonales.
Terapia hormonal	Controla y disminuye los síntomas de moderados a severos, por otro lado se ha demostrado que su administración incrementa el riesgo de desarrollar cáncer de mama y está directamente relacionado con la dosis y el tiempo de uso.
Protocolo HABITS	Estudio controlado, doblemente ciego.
Tibolona	La tibolona es una opción disponible para el tratamiento de los síntomas de la menopausia, y los datos a corto plazo indican su eficacia.
Tratamiento sistémico adyuvante	Se le llama adyuvancia a todo tratamiento antineoplásico administrado después de un manejo quirúrgico; sus objetivos son prolongar el periodo libre de enfermedad, reducir las recurrencias locales y sistémicas y aumentar la supervivencia global.
Hormonoterapia adyuvante	La hormonoterapia adyuvante debe indicarse al menos por 5 años a todas las pacientes con receptores hormonales positivos, para prevenir enfermedad metastásica, recurrencia locoregional y tumores contralaterales.
Hormonoterapia adyuvante extendida	Dirigida a paciente con alto riesgo para recurrencia tardía: tumores menores a 2 cm más factores de riesgo asociados como ganglios positivos, tumores alto grado, pacientes premenopáusicas, alto riesgo para segundo cáncer primario.
Quimioterapia adyuvante	Destruye las células cancerosas al detener la capacidad de multiplicarse de las células. Su quimioterapia puede durar de 3 a 6 meses o más.

Antraciclinas	Antibiótico citostático obtenido a partir de cultivos de microorganismos Streptomyces. Las antraciclinas son el grupo más utilizado: adriamicina (doxorubicina) y su derivado, epirubicina; daunomicina (daunorrubicina) y su derivado, idarrubicina; y mitoxantrona.
Capecitabina Adyuvante	Debe considerarse en pacientes con enfermedad triple negativa que no alcanzan respuesta patológica completa a la neoadyuvancia.
Bisfosfonatos	Permiten mejorar los resultados en la salud ósea al reducir la osteopenia u osteoporosis secundarias al tratamiento sistémico.
Paclitaxel	El paclitaxel es un fármaco utilizado para el tratamiento del cáncer. Fue descubierto por el Research Triangle Institute en 1968.
Taxanos	Los taxanos son uno de los tratamientos más utilizados en el manejo del cáncer de mama desde hace décadas, tanto en fases iniciales como avanzadas. Su uso se extiende a casi todos los tumores sólidos y han mejorado el pronóstico de tumores como el de pulmón, próstata y páncreas.
Terapia Sistémica	Uso de medicamentos para destruir las células cancerosas.
EndoPredict	Prueba de 12 genes que se puede utilizar en pacientes con tumores RH positivos, HER-2 negativo T1 o T2 y ganglios negativos.
MammaPrint	Prueba de 70 genes que tiene utilidad pronóstica en la que se genera un resultado como bajo o alto riesgo genómico.
Oncotype DX	Prueba de 21 genes con valor pronóstico y predictivo, con una amplia validación en el que se genera un puntaje de recurrencia de acuerdo a la expresión de cada uno de los genes.
PAM50 (Prosigna)	Se puede utilizar en pacientes con tumores RH positivos, HER-2 negativo, T1 o T2, ganglios negativos.
Terapias Blanco	Tratamiento contra el cáncer en el que se utilizan fármacos que actúan sobre genes y proteínas específicos que intervienen en el crecimiento y la supervivencia de las células cancerosas.

Pertuzumab	Este anticuerpo monoclonal se puede administrar con trastuzumab y quimioterapia, ya sea antes o después de la cirugía para tratar el cáncer de seno en etapa inicial, o para tratar el cáncer de seno avanzado.
Trastuzumab	Anticuerpo monoclonal humanizado semejante a los anticuerpos que produce el organismo de forma natural para protegerse de las infecciones por virus y bacterias.
Radioterapia	Tratamiento con rayos (o partículas) de alta energía que destruyen las células cancerosas. Es un estándar en pacientes con cáncer de mama localmente avanzado.
Radioterapia Acelerada Parcial	Es un tratamiento alternativo que solo incluye el blanco de radiación al área que rodea la cavidad de la tumorectomía posterior a la cirugía conservadora.
Radioterapia Hipofraccionada	Se realiza con una planeación 3D. Se recomienda que la cobertura del 95 % del volumen del PTV reciba el 95 % de la dosis y no mayor del 105 % de la dosis de prescripción.
Boost Quirúrgico	En pacientes seleccionadas post cirugía conservadora debe ofrecerse una dosis adicional de radioterapia al lecho quirúrgico debido al riesgo de recurrencia local.
Rehabilitación Física	Orientada a la mejora de la calidad de vida de la paciente tratada por cáncer de seno, su objetivo principal es el recuperar el máximo nivel posible de funcionalidad y aliviar el dolor.
Linfedema	Inflamación en un brazo o una pierna ocasionada por una obstrucción del sistema linfático.
Tétrada de Celso	Tumor y rubor con calor y dolor en el brazo.
Terapia Descongestionante Compleja TDC	Esta terapia es suave, no invasiva y en la mayoría de los casos devuelve a la paciente el control sobre su linfedema y la reincorpora a una vida funcional.
Vendaje neuromuscular	Técnica de vendaje realizada con una venda elástica, transpirable y ligera, de hilo de algodón trenzado en forma de cinta, llamada tape kinesiológico.
Drenaje Linfático Manual DLM	Consiste en la aplicación de una tracción sobre la piel de forma lenta y rítmica para conseguir «abrir» las vías linfáticas. De esta forma, se consigue drenar los líquidos excedentes que bañan las células y evacuar

	los desechos que provienen del metabolismo celular.
Presoterapia Secuencial PTS	Realiza una compresión externa que aumenta el flujo circulatorio por diferencia de presión, facilitando el retorno venoso y linfático.
Tratamiento Físico Combinado TFCL	Su objetivo se centra en mantener la autosuficiencia el mayor tiempo posible, preservando la movilidad y la fuerza muscular y disminuyendo notablemente el dolor.
Biopsia	Procedimiento diagnóstico que consiste en la extracción de una muestra total o parcial de tejido para ser examinada al microscopio por un médico anatomopatólogo.
Biopsia con aguja de Corte	Método ideal de diagnóstico de lesiones no palpables; se realiza bajo anestesia local siendo un procedimiento bien tolerado, ambulatorio y con mínimas complicaciones.
Biopsia aspiración por con aguja fina BAAF	El médico utiliza una aguja hueca muy fina adherida a una jeringa para extraer (aspirar) una pequeña cantidad de tejido o líquido de la región que causa sospecha.
Biopsia Quirúrgica	Se realiza una cirugía para extirpar toda o parte de una masa para examinarla y saber si contiene células cancerosas.
Ecografía	También llamada ultrasonografía o ecosonografía, es un procedimiento de diagnóstico usado en los hospitales y clínicas que emplea el ultrasonido para crear imágenes bidimensionales o tridimensionales.
Estereotáxica	Intervención quirúrgica mínimamente invasiva que utiliza un sistema de coordenadas tridimensional para localizar pequeñas estructuras dentro del cuerpo y para realizar acciones tales como ablación (extirpación), biopsia, lesión, inyección, estimulación, implantación de dispositivos, radiocirugía etc.
Diagnóstico Temprano	Es la búsqueda de una enfermedad antes de que la persona presente síntomas.
Estudios de imagen	Permiten detectar, caracterizar, evaluar la extensión de la enfermedad y dar seguimiento a las lesiones mamarias.

Mastografía	Es un estudio de rayos “X” que se recomienda a mujeres de 40 a 69 años de edad, sin signos, ni síntomas de cáncer (asintomáticas) y tiene como propósito detectar anomalías en las mamas, que no se pueden percibir por la observación o la palpación.
Mastografía diagnóstica	Se efectúa en caso de una mastografía de detección anormal.
Mastografía digital	Emplea un detector digital. La adquisición, el procesamiento y la visualización de la imagen se manejan de forma independiente, lo cual representa una mayor ventaja en relación con el sistema análogo,
Biopsia estereotáctica con tomosíntesis	Cuando se encuentra una distorsión de la arquitectura, sin traducción en ultrasonido, la biopsia debe guiarse con sistema corte-aspirado y estereotáctica integrada con tomosíntesis, ya que generalmente se trata de carcinomas invasores.
Mastografía Contrastada	Estudio funcional que combina la mastografía convencional con administración intravenosa de medio de contraste. Existen dos modalidades temporal y dual. Su finalidad es detectar tumores de pequeño tamaño
Mastografía Sintetizada	Es una técnica que consiste en obtener imágenes tridimensionales y a partir de ellas las bidimensionales, por lo cual se reduce la dosis de radiación en 39 %.
Mastografía con tomosíntesis tridimensional o	Se obtienen múltiples imágenes desde diferentes ángulos, mejora la detección de cáncer de mama en un 27 % y disminuye el número de llamadas en programas de tamizaje en un 17.2 %.
Sistema de detección asistida por computadora CAD	Sistemas que orientan en la detección de tumores en una imagen médica, al funcionar como un segundo lector.
Telemastografía	Es una herramienta que permite el envío de las imágenes para su interpretación o consulta a distancia.



Resonancia Magnética	Método complementario a la mastografía y el ultrasonido. No utiliza radiación ionizante y proporciona información morfológica y funcional, a través de la administración endovenosa de medio de contraste paramagnético (gadolinio), es necesario contar con un resonador de por lo menos 1.5 teslas y antena dedicada a la glándula mamaria.
Resonancia Magnética Abdominal	Examen imagenológico que utiliza imanes y ondas de radio potentes. Estas ondas crean imágenes del interior de la zona abdominal. No emplea radiación (rayos X).
Estudios Moleculares de la Mama	Son estudios de imagen, no solo morfológicos sino funcionales que evalúan la actividad molecular de los tumores malignos, al inyectar mediante vía endovenosa un radiotrazador, generalmente 18-fluorodesoxiglucosa.
Tomografía por emisión de positrones PET	Utiliza pequeñas cantidades de materiales radioactivos denominados radiosondas o radiofármacos, una cámara especial y una computadora para evaluar las funciones de tejidos y órganos.
Mastografía por emisión de positrones PEM	Es un estudio de medicina nuclear molecular que utiliza radios trazadores administrados de forma intravenosa, adquiriendo imágenes para localización de la lesión tumoral, así como la valoración de su metabolismo.
PET-CT	Combina tomografía computada (TC) y medicina nuclear, con la localización precisa del tumor primario, así como de las metástasis a distancia, evaluación de respuesta a tratamiento, seguimiento y re-etapificación.
Medicina Nuclear	Especialidad de la medicina en la que se utilizan radiotrazadores o radiofármacos para el diagnóstico y tratamiento de enfermedades. Estos radiofármacos se aplican dentro del organismo humano por diversas vías.
Gammagrafía ósea	Ayuda a diagnosticar y evaluar una variedad de enfermedades y condiciones de los huesos utilizando pequeñas cantidades de materiales radiactivos llamados radiosondas que se inyectan en el torrente sanguíneo.



Tomografía Tórax	Examen imagenológico que utiliza imanes y ondas de radio potentes. Estas ondas crean imágenes del interior de la zona torácica, en caso de síntomas pulmonares.
Estudios de extensión para estadificación inicial	La estadificación del cáncer es el proceso de descubrir cuánto cáncer hay en el cuerpo de una persona y cuánto se ha propagado. Es la manera en la que el médico determina la etapa (estadio) del cáncer de una persona.
Ultrasonido Mamario	Valiosa herramienta complementaria de la mastografía diagnóstica. Se requieren equipos de alta resolución, además de experiencia y conocimiento de la anatomía, la patología de la glándula mamaria y su evaluación por US; debe realizarse con transductor lineal de alta frecuencia, banda ancha y zona focal variable (ideal entre 10 y 23MHz).
Estudios de tamizaje	Permite tomar medidas antes de que se presenten los síntomas. La mayoría de estas enfermedades son muy poco frecuentes, pero son tratables si se detectan a tiempo.
Autoexamen Mamario	Chequeo que una mujer se realiza en casa para buscar cambios o problemas en el tejido mamario.
Examen Clínico Mamario	Evaluaciones médicas para el control del cáncer de mama.
Cáncer de mama en el hombre	El cáncer mamario, en el hombre, representa menos de 1 % del total de los casos de cáncer de mama.
Carcinoma	Tumor maligno que se forma a partir del tejido epitelial de los órganos.
Carcinoma Ductal in Situ CDIS	Es un grupo heterogéneo de neoplasias, caracterizado por la presencia de células epiteliales malignas que crecen dentro de los conductos mamaros, sin rebasar la membrana basal, y se identifican por microscopia de luz.
Ganglio linfático centinela	Es el primer ganglio linfático adonde es más probable que se diseminen las células cancerosas del tumor primario.
Carcinoma lobulillar in situ CLIS	Es una lesión poco frecuente, en la que el diagnóstico histológico y diferencial con la hiperplasia atípica, requiere la intervención de patólogos expertos.

Carcinoma ductal invasor	Es el tipo más común de cáncer de seno.
COVID-19	La pandemia por SARS-CoV-2, decretada por la OMS en marzo del 2020, se convirtió en una emergencia de salud pública global que representó un reto para combinar la continuidad de la atención de los pacientes oncológicos, con la seguridad de pacientes y trabajadores.
Cáncer de mama inflamatorio	Cáncer poco frecuente y de rápida expansión que provoca que los pechos estén rojos, inflamados y sensibles.
Cáncer Mama Metastásico	Enfermedad heterogénea, hasta el momento incurable, con manifestaciones clínicas variables y cuyo tratamiento depende del sitio y el número de las metástasis, las características de la paciente, el inmunofenotipo tumoral y la sensibilidad o la resistencia a los tratamientos médicos oncológicos previos. <sup>1</sup>
Receptores Hormonales Negativos	Son más comunes en mujeres que aún no han tenido menopausia. Las células del cáncer de mama que son triples negativas no contienen receptores de estrógeno ni de progesterona. Tampoco producen exceso de proteína HER2.
HER 2 Neu Positivo	Estos cánceres tienden a crecer y propagarse más rápido que otros tipos de cáncer de mama, pero responden al tratamiento con medicamentos que tienen como blanco a la proteína HER2.
Receptores Hormonales Positivos	Algunos cánceres de mama tienen receptores que se unen a las hormonas, el estrógeno y la progesterona a medida que circulan por el cuerpo. Estas hormonas "alimentan" la célula y la ayudan a crecer. Si su tumor tiene receptores hormonales, se llama receptor hormonal positivo o HR+.
HER 2 Neu Negativo	Se produce cuando las células cancerosas no contienen niveles altos de la proteína HER2.
Recurrente Triple Negativo	Las células de este cáncer no contienen receptores de estrógeno ni de progesterona. Tampoco producen exceso de la proteína HER2.

Bevacizumab	Anticuerpo monoclonal humanizado. Un anticuerpo monoclonal es una proteína que reconoce y se une de forma específica y única a otra proteína del organismo
BRCA	Genes que protegen a las células fabricando proteínas que ayudan a prevenir la formación de tumores.
Olaparib	Es un medicamento indicado para el tratamiento del cáncer de ovario, cáncer de trompa de Falopio o cáncer de mama.
Sales de Platino	Clase de fármacos quimioterapéuticos
Embarazo	Se define como cáncer asociado al embarazo aquel que se diagnostica durante el periodo de gestación, y hasta el primer año posterior a la terminación del embarazo
Enfermedad metastásica	Se refiere al cáncer que se ha propagado del lugar donde se originó hacia otras partes del cuerpo.
Metástasis Cerebrales	Asociadas a tumores con receptores hormonales negativos, HER2 positivo, pacientes premenopáusicas y con enfermedad metastásica en pulmón y/o hígado.
Metástasis Hepáticas	Se refiere a cáncer que se ha diseminado al hígado desde otra parte del cuerpo. La metástasis hepática no es lo mismo que el cáncer que comienza en el hígado, el cual se llama carcinoma hepatocelular.
Metástasis Pulmonares	Las metástasis pulmonares son tumores metastásicos a los pulmones que se desarrollan en otras localizaciones (tumor primario) y se propagan a través del torrente sanguíneo a los pulmones.
Metástasis óseas	Las metástasis óseas son diferentes de un tumor óseo primario (es decir, un tumor que comenzó en el hueso).
Radioterapia Estereotáxica Extracraneal SBRT	Técnica de irradiación de alta precisión que permite irradiar con dosis ablativas, lesiones localizadas fuera del cráneo, como son tumores o metástasis pulmonares, hepáticas, suprarrenales, vertebrales, tumores de páncreas, de próstata o tumores cabeza y cuello.

Resección Paliativa	La extirpación radical del tumor primario dejando enfermedad metastásica residual a distancia.
Resección Tumor Primario	La extirpación radical del tumor primario basado en recomendaciones de estudios retrospectivos.
Factores de Riesgo	Cualquier característica o circunstancia detectable de una persona o grupo de personas que se sabe asociada con la probabilidad de estar especialmente expuesta a desarrollar o padecer un proceso mórbido, sus características se asocian a un cierto tipo de daño a la salud.
Factores de riesgo biológicos.	Posible exposición a microorganismos que puedan dar lugar a enfermedades, motivada por la actividad laboral. Su transmisión puede ser por vía respiratoria, digestiva, sanguínea, piel o mucosas.
Disfunción Cognitiva	Las causas de esta compleja toxicidad que se presenta a mediano y largo plazos son hasta el momento poco claras. La incidencia de daño cognitivo secundario a quimioterapia es de 20 % a 30 %.
Estilo de vida	El estilo de vida entendido como un patrón de comportamiento, está determinado por factores como: el entorno social y laboral, el ambiente, la alimentación, la recreación, el nivel educativo, el sexo, actividad física, etc.
Fatiga	Se denomina así a la sensación persistente de cansancio asociada de manera no proporcional con actividades físicas. Se presenta hasta en 80 % de las pacientes tratadas con quimioterapia y persiste por 6 a 12 meses después de la finalización de la misma en 30 % de los casos.
Insuficiencia ovárica por quimioterapia	Las dosis altas pueden destruir algunos o todos los óvulos en los ovarios y podrían causar infertilidad o una menopausia precoz. La mayoría de las mujeres que reciben radiación pélvica perderán su fertilidad.
Neuropatía	La neuropatía es una complicación muy frecuente en pacientes que reciben tratamiento con taxanos. La incidencia es de 13 % a 27 % y varía según el tipo y la frecuencia del taxano utilizado.

Duloxetina	Antidepresivo inhibidor de la recaptación de serotonina y noradrenalina utilizado para el tratamiento de la depresión mayor, así como el dolor asociado con la neuropatía diabética y la fibromialgia.
Gabapentina	La gabapentina es un medicamento originalmente desarrollado para el tratamiento de la epilepsia. Posteriormente se empezó a utilizar para el tratamiento del dolor, especialmente el de origen neuropático.
Pregabalina	Fármaco antiepiléptico y analgésico usado en el dolor neuropático periférico, y como terapia añadida en las crisis parciales convulsivas con o sin generalización en adultos.
Reproductivos	Es la posibilidad para mujeres en edad reproductiva de sufrir daños en su salud o en la de su futuro hijo si decide embarazarse.
Síntomas Menopáusicos	La prevalencia de los síntomas climatéricos inducidos por la quimioterapia y el tratamiento hormonal (bochornos y sudoraciones nocturnas, resequedad vaginal y atrofia, incontinencia, dispareunia, insomnio, irritabilidad, artralgias, fatiga) varía de acuerdo con la edad, el tipo de tratamiento y el número de ciclos de quimioterapia administrados.
Síntomas Menopáusicos	La prevalencia de los síntomas climatéricos inducidos por la quimioterapia y el tratamiento hormonal (bochornos y sudoraciones nocturnas, resequedad vaginal y atrofia, incontinencia, dispareunia, insomnio, irritabilidad, artralgias, fatiga) varía de acuerdo con la edad, el tipo de tratamiento y el número de ciclos de quimioterapia administrados.
Toxicidad	La toxicidad es la capacidad de una sustancia química de producir efectos perjudiciales sobre un ser vivo, al entrar en contacto con él.
Cardiotoxicidad	Condición en la que hay daño al músculo del corazón.

Adriamicina	Fármaco ampliamente utilizado en la quimioterapia del cáncer. Es un antibiótico de la familia de las antraciclinas, con una estructura muy similar a la daunomicina, siendo al igual que ésta un intercalante del ADN.
Epirubicina	Fármaco de antraciclina utilizado para la quimioterapia. Se puede usar en combinación con otros medicamentos para tratar el cáncer de mama en pacientes que se han sometido a una cirugía para extirpar el tumor.
Leucemia	Cáncer de las células primitivas productoras de sangre. Con mayor frecuencia, la leucemia es un cáncer de los glóbulos blancos, pero algunas leucemias comienzan en otros tipos de células sanguíneas.
Leucemia Mielocítica Aguda	Tipo de cáncer hematológico y de la médula ósea con exceso de glóbulos blancos inmaduros.
Síndrome mielodisplásico	Conjunto de trastornos ocasionados por la interrupción en la producción de células sanguíneas.
Tratamiento de enfermedades	Conjunto de medios (higiénicos, dietéticos, farmacológicos, quirúrgicos o físicos) que se ponen en práctica para curar o aliviar una enfermedad. También se dice de un medio especial de cura.
Histologías poco frecuentes	Aquellas que afectan a un número pequeño de personas en comparación con la población general y que, por su rareza, plantean cuestiones específicas.
Tumor Phyllodes	Neoplasia fibroepitelial, que representa el 1 % de los tumores de mama.
Integración Cuidados Soporte Paliativos	Asistencia activa, holística, de personas de todas las edades con sufrimiento severo, relacionado con la salud debido a una enfermedad grave, y especialmente de quienes están cerca del final de la vida.
Lactancia	Forma de alimentación que comienza en el nacimiento con leche producida en el seno materno.

Mujeres jóvenes	Mujer con edad igual o menor de 40 años. Esta delimitación se basa en las diferencias observadas con respecto a los factores de riesgo, las características tumorales y los desenlaces clínicos, así como en los intereses particulares para este grupo de edad: fertilidad, autoimagen, percepción de la calidad de vida y objetivos personales.
Revista Indexada	Publicación periódica de investigación que denota alta calidad y ha sido listada en alguna base de datos/índice/repertorio de consulta mundial.
Base de Datos	Conjunto de datos estructurados que pertenecen a un mismo contexto.
IMBIOMED	índice virtual de publicaciones médicas y científicas, de amplia implantación en Latinoamérica y España.
Latindex	Sistema de información académica, sin fines de lucro y de consulta gratuita.
Lilacs	Control bibliográfico y la diseminación de la literatura científico técnica latinoamericana y del Caribe en el área de la salud, ausentes de las bases de datos internacionales.
Medigraphic	Empresa editorial especializada en el ramo biomédico y científico.
Refereada	Hace referencia.
Tipo Documento	Formato en el que se encuentra la publicación, PDF/HTML
Articulo	Trabajo de investigación o comunicación científica publicado en alguna revista especializada.
Revista	Publicación periódica en la que se intenta recoger el progreso de la ciencia, entre otras cosas incluyendo informes sobre las nuevas investigaciones.
Tesis	Inicio de un texto argumentativo, una afirmación cuya veracidad ha sido argumentada, demostrada o justificada de alguna manera.



Este listado de conceptos se pretende utilizar más adelante como trabajo a futuro en caso de que una persona no conozca el significado de algún concepto, podrá en el sistema consultar este significado a manera de retroalimentación.

### 4.1.3 Visualización de la Taxonomía implementada

Con apoyo del programa Protégé (Stanford University, 2016-2020) que es un editor de ontologías, se implementó la estructura de la taxonomía, como se puede ver en la Figura 7, el primer paso en realizar fue declarar las entidades (conceptos) de acuerdo con el diagrama de flujo de la Figura 6, definiendo quienes serían las entidades principales (facetas) y posteriormente sus entidades hijos.

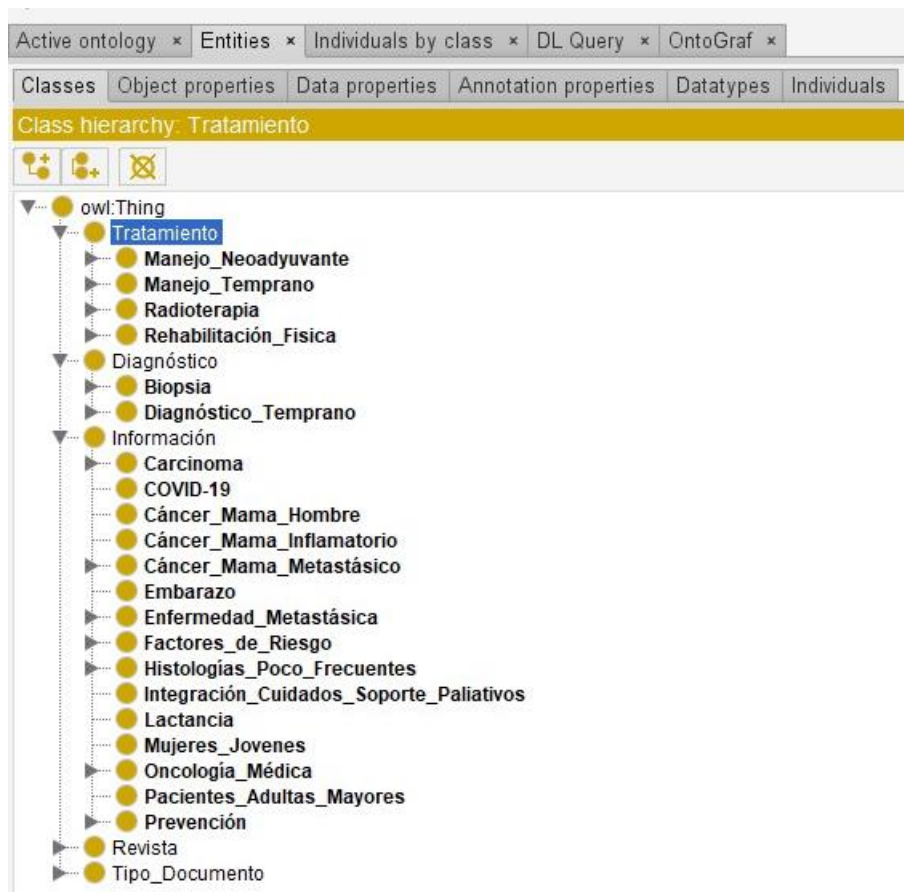


Figura 7 - Definición de entidades en Protégé



Una vez definidas las entidades, se puede obtener una visualización gráfica de la taxonomía por parte de la herramienta OntoGraf la cual provee apoyo visual de cada una de las entidades y sus hijos, se puede obtener este detalle en vista radial, o de árbol, esto con el fin de apreciar los niveles y las uniones entre cada uno de los conceptos de otra manera. En la Figura 8, se puede observar la estructura de la taxonomía en forma de árbol horizontal. Se tienen un total de 166 conceptos pertenecientes a la taxonomía por lo que en esta figura se notan bastantes términos y es difícil apreciar con claridad las leyendas de los conceptos.

Cabe señalar que sólo se usa protégé con el propósito de ilustrar la taxonomía mediante OntoGraf, pero no se utiliza protégé para la implementación del sistema de recomendación objeto de este trabajo de investigación.

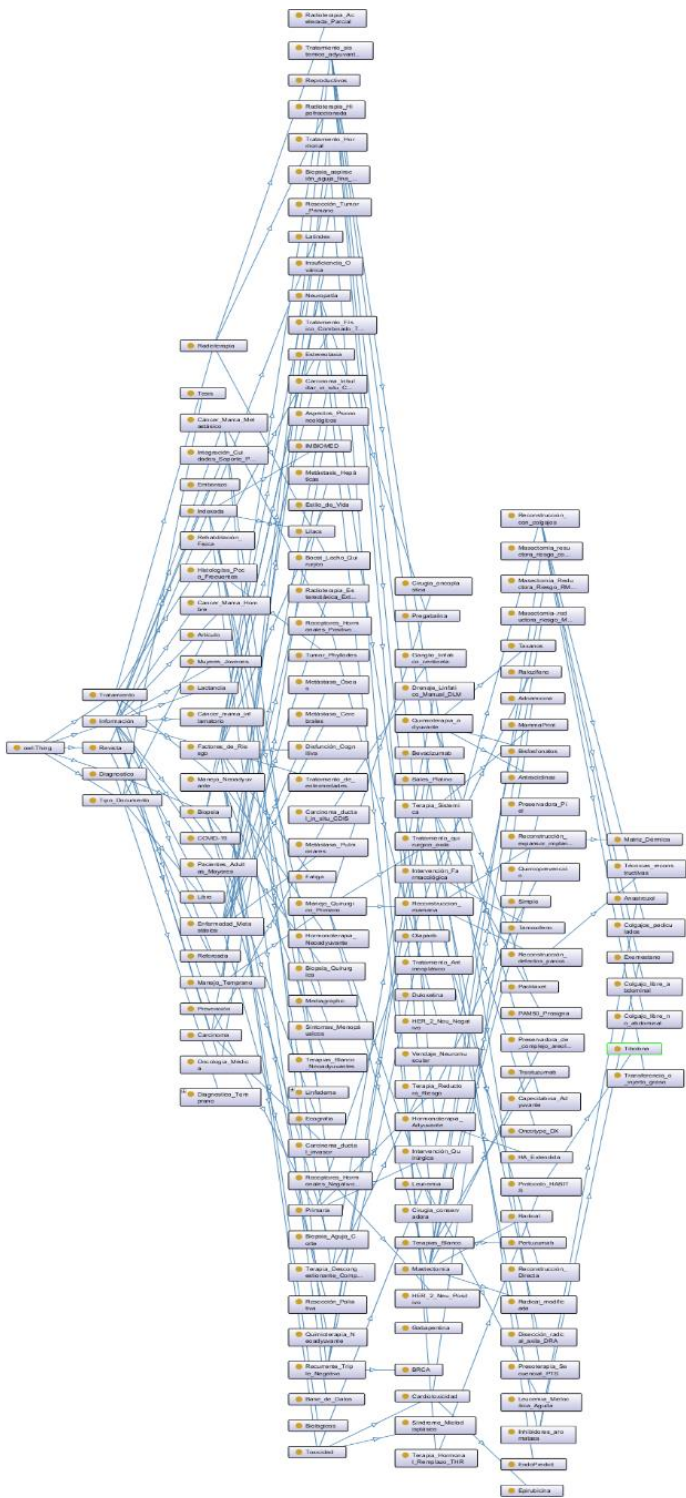


Figura 8 - Taxonomía horizontal

Para comprender y visualizar la estructura en la Figura 9, se puede apreciar un gráfico radial de la taxonomía, donde se puede ver como a partir de la raíz se relacionan *Tratamiento*, *Información*, *Diagnóstico*, *Tipo\_Documento* y *Revista* que son las facetas de nuestra taxonomía, las cuales a su vez tendrán relación con uno o varios términos de acuerdo con el significado de la faceta y cada termino a su vez puede tener relacionado más de un término, teniendo así un funcionamiento de taxonomía facetada, en caso de algunos conceptos se puede apreciar un icono de +, lo que quiere decir que tiene relacionados otros conceptos como hijos y esos hijos pueden tener más términos relacionados de manera anidada como se puede apreciar en la Figura 8.

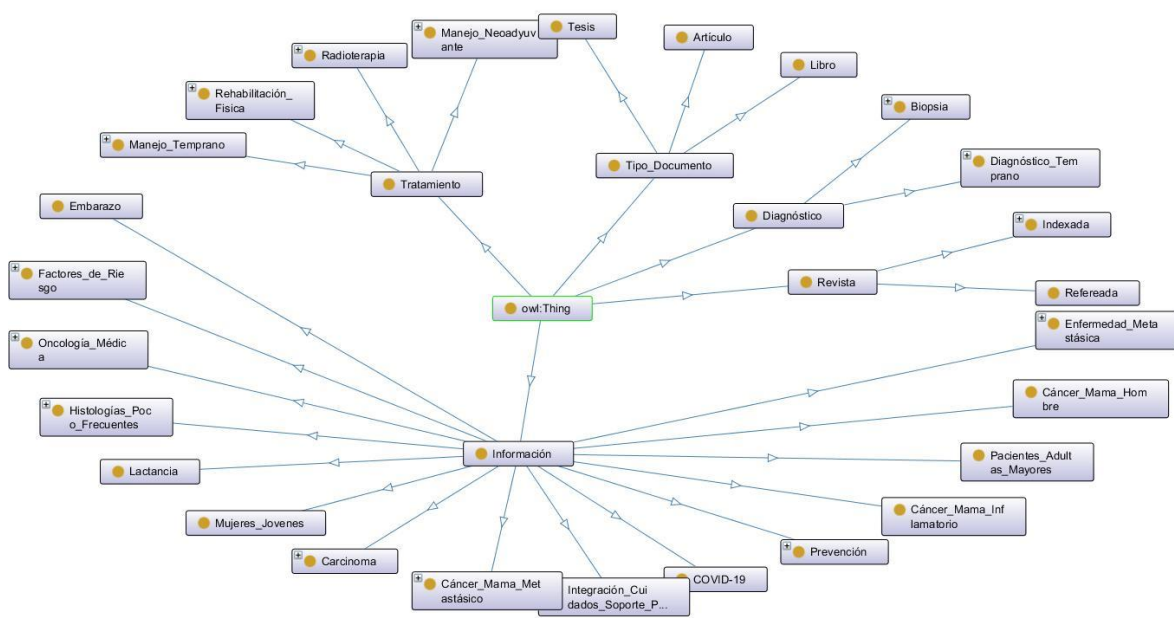


Figura 9 - Taxonomía Radial, facetas y algunos conceptos

Como ya se mencionó en la especificación de requisitos, esta taxonomía se representará por medio de una base de datos para poder indizar los documentos que se obtendrán con el proceso de extracción, de igual manera la base de datos será de uso para la gestión de usuarios y otras configuraciones del Sistema de Recomendación.

En la Figura 10, se presentan todos los conceptos de la taxonomía de manera radial, partiendo del centro y extendiendo todos los conceptos relacionados con sus facetas y conceptos, teniendo en color verde el concepto más profundo de la taxonomía.

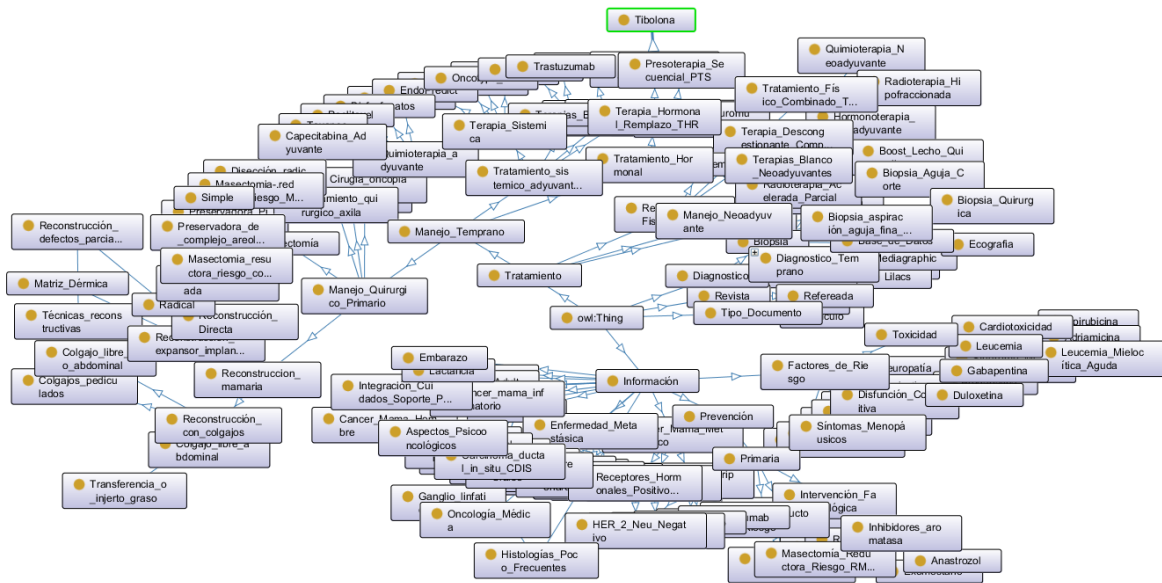


Figura 10 - Conceptos completos de la Taxonomía

## Capítulo 5

### 5.1 Arquitectura del Sistema de Recomendación

Teniendo hasta el momento un único usuario se explica a continuación las actividades que podrá desempeñar dentro del sistema. Una vez que el usuario haya iniciado sesión este podrá acceder a un panel principal o Dashboard, donde podrá en primera instancia visualizar la taxonomía e información relevante. Podrá realizar búsquedas con el apoyo del buscador que se encuentra en el Dashboard, y ahí mismo el sistema le mostrará los resultados más relevantes en relación con el criterio de búsqueda que realizó el usuario.

La implementación del Sistema de Recomendación se basa en el modelo Cliente/Servidor, cuya funcionalidad consiste en que la información del sistema se encuentren separados de los contenidos del sistema, a su vez, están separados con toda claridad del aspecto y la sensación de interfaz; Por ejemplo, un usuario, con acceso al sistema desde un dispositivo móvil o PC, ejecuta una acción dentro del sitio, la cual se traduce en una petición al servidor, este ejecuta una consulta a la Base de Datos y dependiendo del resultado de la consulta, el servidor posteriormente devolverá una respuesta de error o confirmación para la petición del cliente (usuario), una explicación grafica se puede observar en la Figura 11.

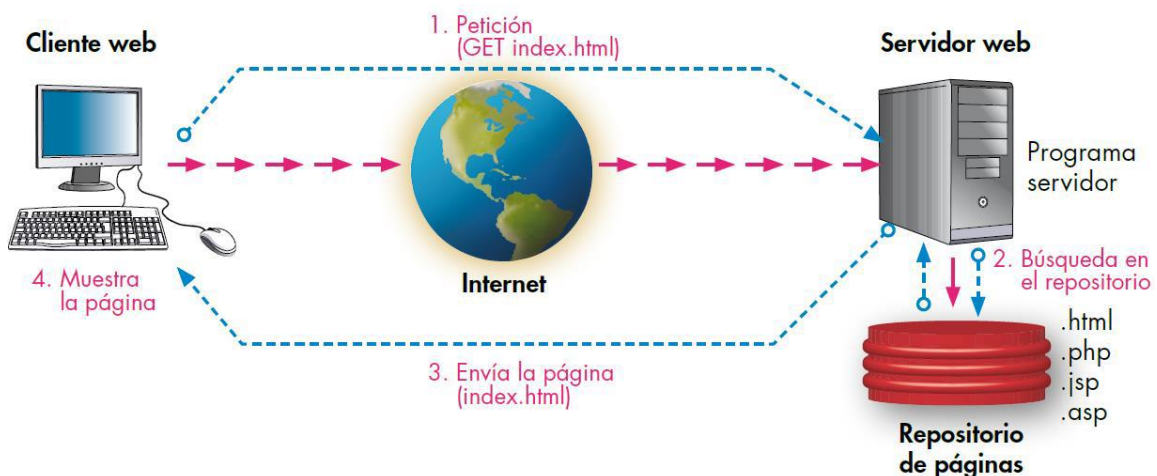
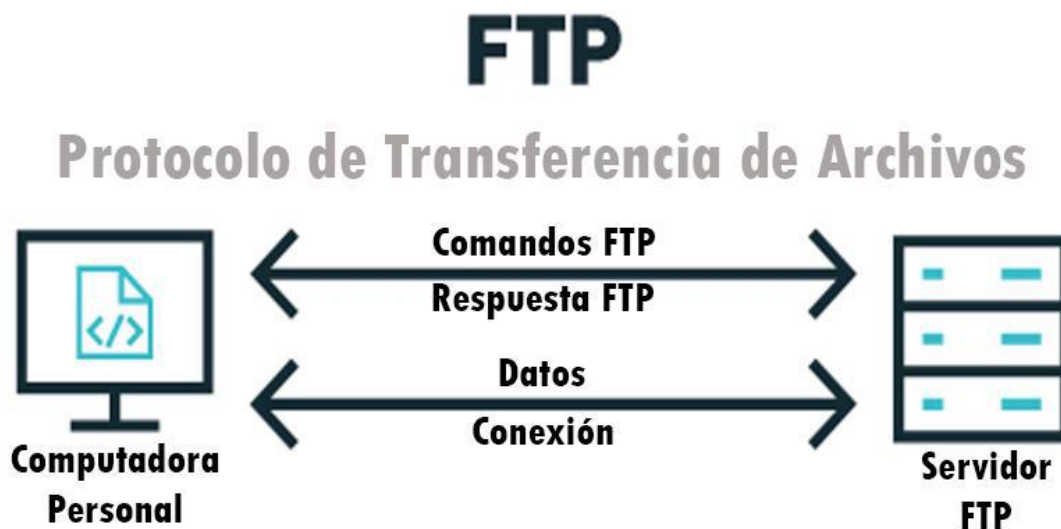


Figura 11 - Ejemplo de modelo Cliente-Servidor

Para almacenar los componentes del sistema de recomendación ya sean documentos, o imágenes, se puede utilizar un servidor FTP. En la Figura 12, se puede visualizar la similitud que tiene con el modelo cliente-servidor. Se requiere de un modo de acceso, que en este caso será “FTP”, para indicar el protocolo por el cual se administrará la conexión y manipulación de los archivos del sistema, tomando en cuenta el tipo de usuario.



*Figura 12 - Protocolo de Transferencia de Archivos*

Una vez definida la arquitectura Cliente-Servidor del Sistema de Recomendación, es necesario contar con un Modelo de Datos, el cual será utilizado para apoyar el diseño, la construcción y la gestión de los datos de los usuarios y la información que se obtendrá y que serán almacenados en la base de datos.

### 5.1.1 Modelo de datos

Un Sistema de Recomendación como cualquier otro sistema, permite a los usuarios manipular datos a través de su interfaz, por ende, es importante diseñar un modelo de datos, para así lograr tener una buena consistencia en relación de los datos con el usuario.

Un modelo de datos es una colección de herramientas conceptuales para la descripción de datos, relaciones entre datos, semántica de los datos y restricciones de consistencia, este se divide en el Modelo Entidad-Relación y el Modelo Relacional (Silberschatz, Korth, S. Sudarshan, & Sudarshan, 2002).

El modelo entidad-relación (E-R) se basa en la percepción del mundo real, lo cual consiste en una colección de objetos básicos, tratados como entidades, y de relaciones entre estos objetos, siendo un modelo de datos de alto nivel.

El modelo relacional utiliza una colección de tablas para representar tanto los datos como las relaciones entre ellos, siendo un modelo de menor nivel que E-R.

Los diseñadores de bases de datos generan el diseño del esquema de la base modelando primero los datos en alto nivel, usando el modelo E-R, y después traduciéndolo al modelo relacional. Permitiendo que la implementación sea lo más fácil posible.

En la Figura 13, se puede observar el esquema de la representación de una taxonomía por medio de una base de datos, la cual consta de 5 tablas, en la parte inferior se aprecia la tabla Taxonomía que va a almacenar el nombre de la taxonomía, cabe resaltar que se pueden almacenar aquí mismo varias taxonomías. Esta tabla de Taxonomía tiene una relación de 1 a muchos con Facetas, las cuales se relacionan con la tabla términos, con una relación de 1 a muchos, ahí mismo se tiene una relación de uno a uno, un término puede estar relacionado con otro término, estos términos pueden estar relacionados con muchos objetos, por ende, se encuentra la tabla intermedia Termino\_has\_Objeto donde se almacenará el id del Término y el id del Objeto como llaves foráneas para evitar la relación muchos a muchos.

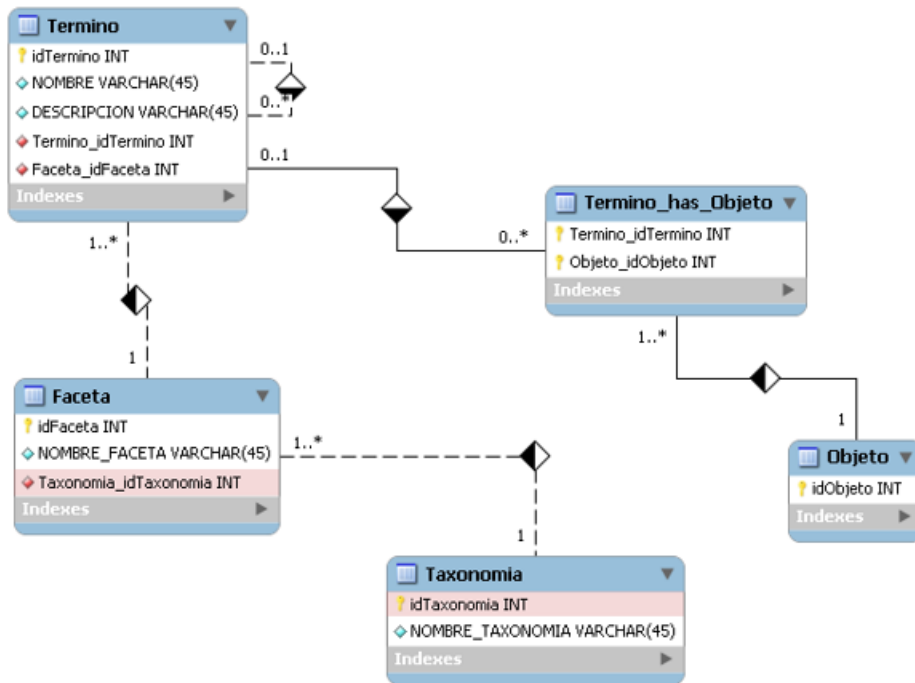


Figura 13 - Esquema Conceptual de la Taxonomía Implementada

Como se puede apreciar en la Figura 14, la arquitectura de la base de datos está compuesta por 8 tablas, que están conectadas unas con otras, para cumplir con una funcionalidad específica y de esta manera se cumpla el modelo relacional.



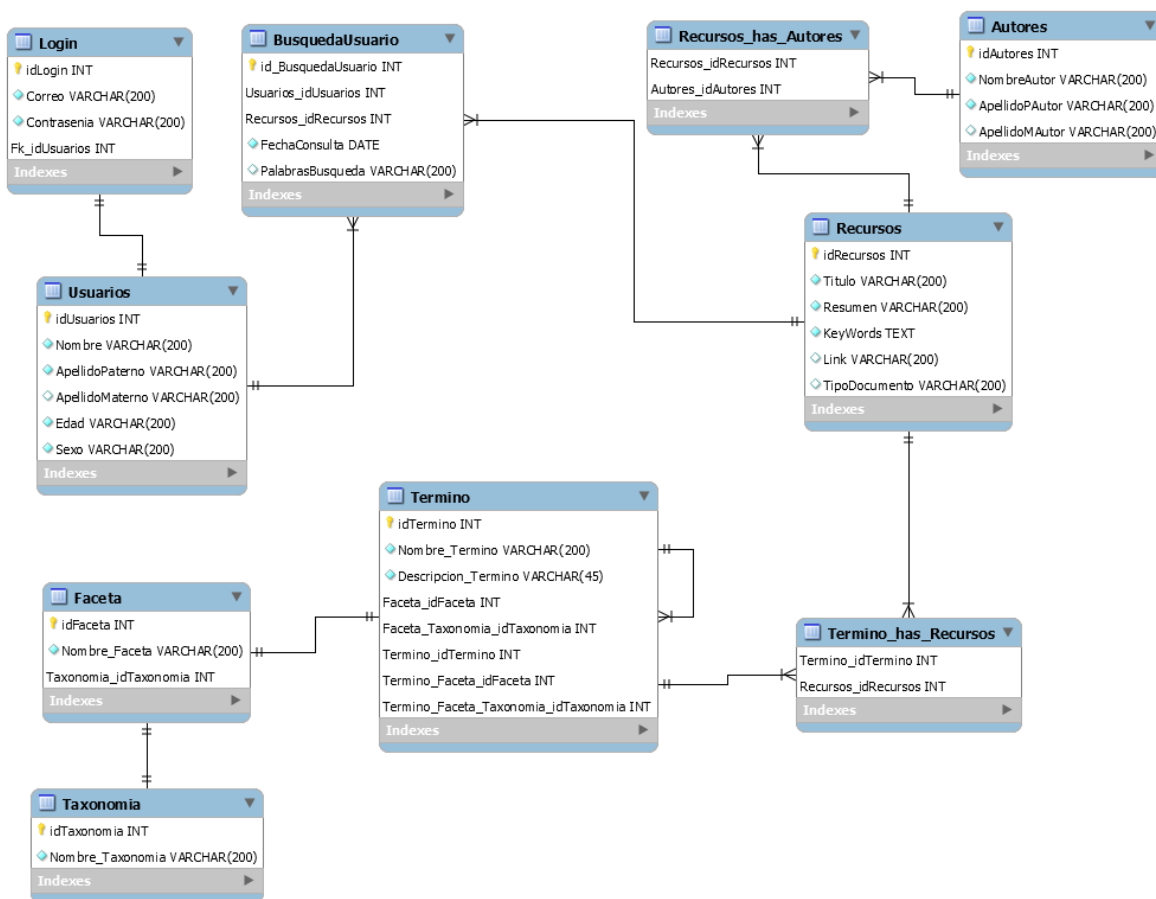


Figura 14 - Modelo Relacional del proyecto

A continuación, en la tabla 2, se describen las entidades que se tienen para la base de datos del proyecto.

**Tabla 2. Entidades de la Base de Datos**

Nombre	Función	Relación
Login	Almacena los datos de los usuarios para iniciar sesión en el sistema.	Usuario
Usuario	Almacena el índice que tiene relación con Login así como detalles de los usuarios.	Login

BusquedaUsuario	Almacena el índice que tiene relación con usuario, así como la relación con recursos, fecha y palabra que un usuario utilizo para una búsqueda.	Usuario, Recursos
Recursos	Almacena la información perteneciente a un recurso.	BusquedaUsuario Recurso_Has_Autores
Autores	Almacena la información de los autores de los recursos	Recurso_Has_Autores
Recurso_Has_Autores	Almacena los índices que tiene relación con Recursos y Autores.	Recurso Autores
Termino	Almacena la información de los conceptos de la Taxonomía, así como los índices que relaciona con Faceta y el mismo.	Termino Faceta Termino_has_Recursos
Termino_has_Recurso	Almacena los índices que relaciona a término y recursos.	Termino Recursos
Faceta	Almacena la información de las facetas de la taxonomía	Termino Taxonomía
Taxonomía	Almacena el nombre de las taxonomías a representar.	Faceta

## 5.1.2 Rastreo y búsqueda de PDF académicos.

Con el paso del tiempo, la web se convirtió en una herramienta muy poderosa dado que almacena infinidad de información. Desarrolladores, especialistas en marketing, e incluso científicos de datos utilizan web scraping para recopilar datos en línea para ayudarles a tomar decisiones valiosas en el desarrollo de sus actividades laborales o de investigación.

### 5.1.2.1 Web Scraping

Web scraping se define como el conjunto de métodos para extraer y obtener información de un sitio web, esta información se recopila y se exporta a un formato útil para que el usuario pueda manipular, ya sea una hoja de cálculo, XML, formato JSON, CSV, BD o incluso una API.

En la Figura 15, se presenta un diagrama de cómo es que opera el web scraping, partamos de que existen infinidad de sitios en la web, todos ellos ricos en información, para obtener esta información se utilizan herramientas o métodos de web scraping, ya sean programas de extracción, scripts desarrollados bajo un lenguaje de programación, o incluso por medio de APIs. Al extraer esta información, para que el usuario pueda utilizarla, se debe almacenar en algún formato de manejo y almacén de información como Excel, archivos XML, CSV o incluso en formato JSON.



Figura 15 - Representación del Web Scraping

En un inicio se realizaron pruebas con herramientas del web scraping con Python y las librerías Selenium y BeautifulSoup. Como se puede apreciar en la Figura 16, se tiene una prueba básica de extracción de información, mediante Selenium (Muthukadan, 2011-2018), se instaló la librería y por medio de sus funciones se accedió al contenido de Google Scholar, obteniendo elementos como el título de la página y otros más importantes como los resultados de las búsqueda realizada en el sitio. Sin embargo, la instalación y configuración de esta librería es bastante compleja ya que utiliza un driver de Google Chrome ya que existen cierta seguridad en los sitios web, que hacen que los *scraps* no puedan obtener información, por esta razón se optó por realizar más pruebas con otras librerías.

```

+ Código + Texto
▶ !pip install selenium

[ ] !apt-get update

▶ !apt install chromium-chromedriver

[ ] !cp /usr/lib/chromium-browser/chromedriver /usr/bin

cp: '/usr/lib/chromium-browser/chromedriver' and '/usr/bin/chromedriver' are the same file

[ ] import sys
sys.path.insert(0, '/usr/lib/chromium-browser/chromedriver')
from selenium import webdriver
from selenium.webdriver.common.by import By
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument('--headless')
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')
wd = webdriver.Chrome('chromedriver', chrome_options=chrome_options)
#wd.get("https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=cancer+ductal&btnG=&og=")
#wd.get("https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=prevenci%C3%B2n+del+cancer+de+mama&btnG=")
wd.get("https://www.inegi.org.mx/programas/mortalidad/#Microdatos")

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:9: DeprecationWarning: use options instead of chrome_options
if __name__ == '__main__':

[ ] wd.title

'Mortalidad'

```

Figura 16 - Pruebas Selenium

Beautiful Soup (Richardson, 2004-2015) es otra poderosa librería que se utiliza para obtener información de los sitios web, en la Figura 17, se puede observar el funcionamiento, desde su sencilla instalación, ya que, a comparación de Selenium, no necesita de un tercero para realizar sus funciones, pudiendo obtener información importante de un sitio web como el título e la página, los contenedores *div* que almacenan la información relevante de un sitio e incluso obtener información a través de los atributos de los elementos que conforman un sitio.

```
[1] !pip install beautifulsoup4
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.7/dist-packages (4.6.3)

[3] soup.title
<title>Página de prueba</title>

[4] div_main = soup.div
div_main['id']
'main'

soup.div.div.div
<div class="links">
<a href="https://pagina1.xyz/">Enlace 1</a>
<a href="https://pagina2.xyz/">Enlace 2</a>
</div>

[6] primer_parrafo = soup.p
texto = primer_parrafo.string
texto
'Este es el primer párrafo'
```

Figura 17 - Pruebas Beautiful Soup

### 5.1.2.2 Desventajas del Web Scraping

A pesar de que pueden ser de gran utilidad las técnicas del *web scraping*, estas no están libres de riesgos legales. Debido a que los desarrolladores crean los sitios para que lo usen personas, así que acceder a la información mediante un *web scraper* estaría infringiendo las condiciones de uso del sitio en cuestión. Dichas acciones son especialmente relevantes cuando se trata de acceder a grandes volúmenes de información que proviene de varias páginas al mismo tiempo o en sucesión rápida, de modo que una persona nunca sería capaz de interactuar con la página.

Si la extracción de información se genera de manera automática, la consulta, el almacenamiento y la evaluación de la información publicada por un sitio web podrían infringir los derechos de propiedad intelectual. Además, de que, si la información obtenida es de carácter personal, almacenarlos y analizarlos sin autorización de las personas afectadas infringe claramente la normativa vigente de protección de datos. Por lo que no está permitido, por ejemplo, *scrapear* perfiles de Facebook para obtener datos personales.

Otra de las desventajas al usar de usar web scraping se centra en que sitios importantes de grandes compañías como Google, Facebook, Yahoo, etc., limitan el acceso a los scrapers, por medio del uso del estándar *robots.txt*.

El cuál como se mencionó al utilizar la librería Selenium, este robots.txt no es más que un archivo de texto que los desarrolladores alojan en el directorio principal de sus sitios web. En este archivo se pueden encontrar configuraciones especiales que establecen cuales o que **scrapers** o **bots** tienen autorización de acceso a determinadas áreas del sitio. Estas configuraciones siempre se aplican a un dominio entero, lo cual impide el funcionamiento correcto del *scrap*. Generalmente al bloquear al *scrap*, en automático como siguiente medida de seguridad se bloquea la ip de la máquina de donde se genera la petición haciendo que no se pueda siquiera tener acceso al sitio.

### 5.1.2.3 Alternativa del web scraping

Aun cuando el web scraping puede ser muy efectivo, no es precisamente el mejor método para obtener información de sitios web. Ya que, existe una alternativa mejor que supera las limitaciones del web scraping, pues muchos desarrolladores publican la información de manera estructurada y en un formato legible para las computadoras. Para poder acceder a este tipo de información se deben utilizar interfaces de programación especiales llamadas Application Programming Interfaces (interfaces de programación de aplicaciones, API por sus siglas en inglés). Las APIs son un conjunto de subrutinas, funciones y procedimientos que ofrece cierta biblioteca para ser utilizada por otro software como una capa de abstracción de información. Estas APIs extraen la información con técnicas de web scraping, pero con los recursos necesarios para evitar los archivos de configuración de los sitios web y el bloqueo de las ip de sus servidores.

### 5.1.2.4 Ventajas de las API

El propietario de un sitio web crea la API para permitir el acceso a la información, por lo que de esta manera se reduce el riesgo de infracciones y el desarrollador puede regular el acceso a la información. Comúnmente para obtener acceso a la información el desarrollador debe solicitar una clave API regulando de manera más precisa las limitaciones y el rendimiento de los sitios.

La API presenta la información en un formato legible para las máquinas, como lo que hace un *scrap*, almacenar la información en un archivo csv, Excel o JSON, para que el desarrollador tenga una mejor manipulación de la información.

En este proyecto para extraer la información se utilizó el API llamada SerpApi, para que posteriormente con la información obtenida, se desarrollaran algoritmos bajo el lenguaje de programación Python y el framework Flask para tratar la información, con el fin de devolver al usuario por medio de una interfaz web la información obtenida en relación con una búsqueda que realice el usuario, tomando la información de sitios como Google académico, retornando datos importantes como título, resumen, citas, un enlace para el acceso al recurso académico (PDF), ya sean públicos o de paga. En caso de que los artículos sean de acceso gratuito, el usuario podrá descargarlos en automático y en dado caso de que algún artículo sea de paga, se redirigirá al usuario al sitio que publicó el recurso para que proceda con la compra del recurso, de esta manera no se infringen derechos de propiedad intelectual ni se viola la normativa de la protección de datos.

## Capítulo 6

### 6.1 Implementación

En este apartado se describen las herramientas de programación que se utilizaron para implementar el Sistema de Recomendación, posteriormente se describen los scripts para la obtención de los recursos por medio del API, así como los apartados por vistas del diseño del sistema orientado a los diferentes tipos de usuario, donde se describen las funcionalidades de acuerdo con el usuario que accede al sistema.

**Python** es un lenguaje de programación utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y *machine learning* (ML). Los desarrolladores comúnmente utilizan Python porque es eficiente y muy fácil de aprender, además de que se puede desplegar en muchas plataformas diferentes (Python Software Foundation, 2001-2022).

**Flask** es un framework de Python que permite crear un programa ejecutable en un servidor web con un número mínimo de líneas de código. Un framework es un conjunto estandarizado de conceptos, prácticas y criterios utilizados para enfocar un tipo de problemática particular que se toma como referencia, para enfrentar y resolver nuevos problemas similares (Pallets, 2010).

**SerpApi** es una API en tiempo real de extracción de datos, esta escanea, extrae y produce de manera automática información procesable de las páginas de resultados de Google y otros motores de búsqueda. Administra todas las partes difíciles del scrap de datos como: proxies, CAPTCHA, rotación de IP, cambios de página, emulación de navegador, alto volumen, tiempos de respuesta rápidos etc (LLC, 2022).

**SQL** es un lenguaje de consulta estructurada, comúnmente utilizado como estándar para la administración de bases de datos. SQL describe conjuntos de datos que ayudan a responder preguntas mediante la sintaxis correcta. Esta sintaxis se basa en la sintaxis del idioma inglés. Ahora bien, una instrucción SQL está conformada por cláusulas. Cada cláusula realiza una función dada una instrucción SQL (Silberschatz, Korth, S. Sudarshan, & Sudarshan, 2002).



**MySQL** es un Sistema de Administración de Base de Datos Relacional (RDBMS), sólido, rápido y flexible. Ideal para la creación y administración de Base de Datos para sitios web. Permite la posibilidad de realizar múltiples consultas de manera rápida (Silberschatz, Korth, S. Sudarshan, & Sudarshan, 2002).

**HTML** es un “Lenguaje de Marcado de Hipertexto” o “Lenguaje de Etiquetas”, es el elemento más básico para la construcción de sitios web, utilizado para crear y representar visualmente una página web (Gauchat, 2012). Se encuentra conformado por etiquetas, estas indican como se presentará la información del documento, también describen la estructura y el contenido de manera textual y a su vez complementa el texto con imágenes, videos, audio, etc.

**HTML5** es un nuevo concepto integrado para el desarrollo de aplicaciones o sistemas web, es más que una simple actualización de HTML, dado que en la actualidad es un área que combina dispositivos móviles, computación en la nube y trabajos en red. Dentro de las características básicas que provee HTML5, se encuentran la estructura, el estilo y la funcionalidad (Gauchat, 2012).

### 6.1.1 Inicio de sesión

En la Figura 18, se presenta la pantalla de inicio de sesión, donde los usuarios, podrán *iniciar sesión* para ingresar al sistema, siendo ésta la función principal de este apartado. Se muestran de igual manera dos enlaces, los cuales son de registro por si aún no se tiene una cuenta, mientras que el otro re-direccionará al usuario a la recuperación de la contraseña en caso de no recordarla.

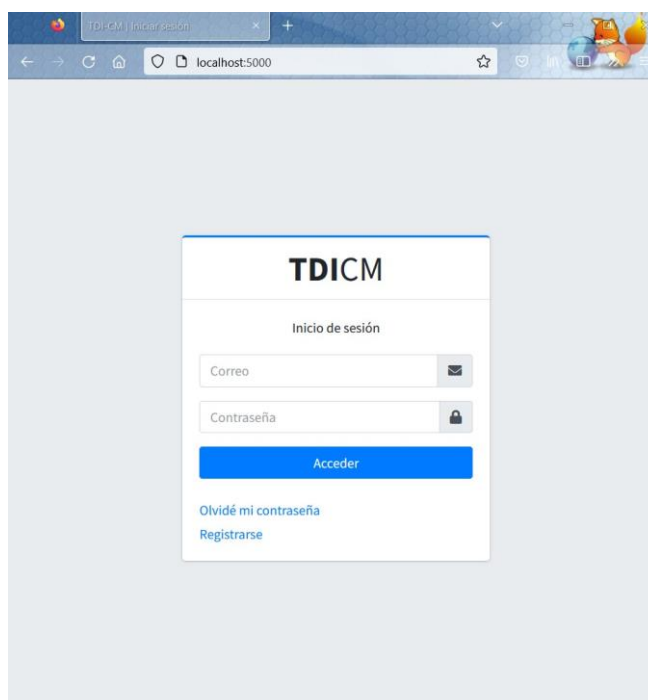


Figura 18 - Inicio de sesión

### 6.1.2 Panel Principal (Dashboard)

En la Figura 19, se muestra el *Panel Principal* o *Dashboard*, del lado izquierdo se tiene de manera lateral el menú de navegación, mostrando la foto del usuario y su nombre, así como un árbol jerárquico que representa la taxonomía desarrollada para este proyecto, de tal manera que el usuario, puede utilizarla para navegar a través de sus conceptos. Del lado derecho en lo que resta de la pantalla se implementó un panel en el que en la parte superior aparece el título "Tratamiento, Diagnóstico e Información Cáncer de mama – TDI" que funciona como enlace, para redirigir al usuario al Dashboard, así como 3 iconos, el primero es una lupa que abrirá el navegador ahí mismo para que el usuario realice sus búsquedas, el siguiente icono con 4 flechas interconectadas sirve para poner en pantalla completa el sitio, el

tercer icono con una flecha apuntando a la derecha, hace referencia al *cierre de sesión*, llevando al usuario al *login* finalizando el acceso al Dashboard. Finalmente, debajo de los iconos, se muestran tres tarjetas con información relevante a el total de artículos indexados, las búsquedas que ha realizado en total ese usuario y los usuarios registrados al momento en el sistema.

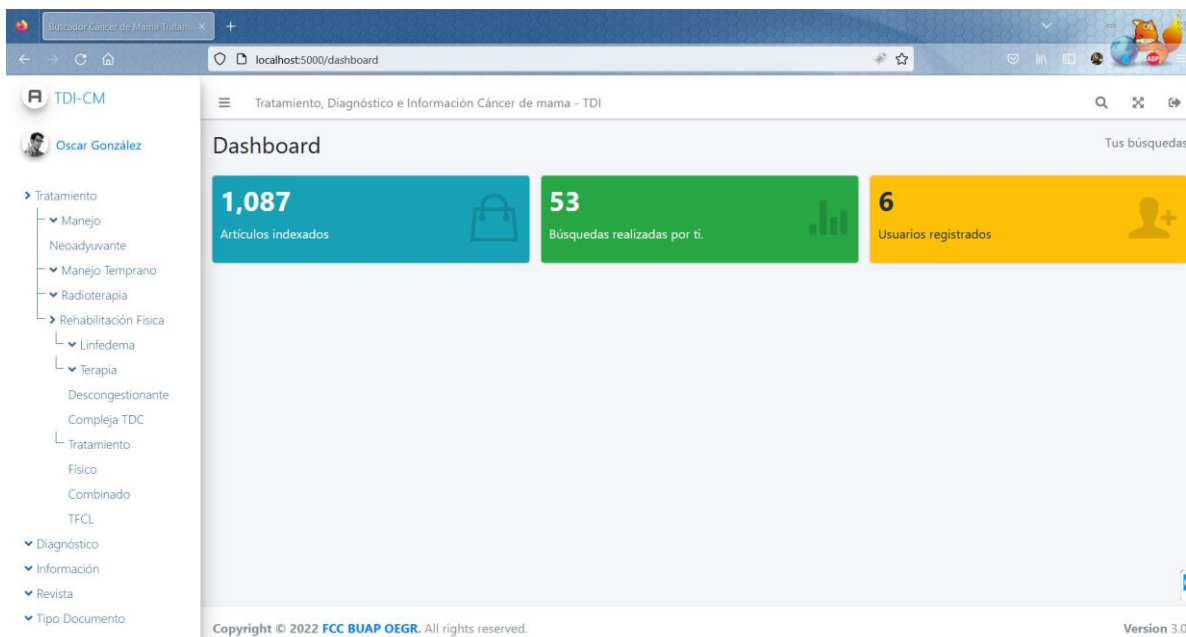


Figura 19 - Panel Principal / Dashboard

### 6.1.3 Búsqueda

Como se puede apreciar en la Figura 20, si un usuario quiere realizar una búsqueda, deberá dar clic en el icono de la lupa, se le desplegará un buscador para que el usuario escriba sus términos o palabras y así el motor de búsqueda le pueda devolver resultados.

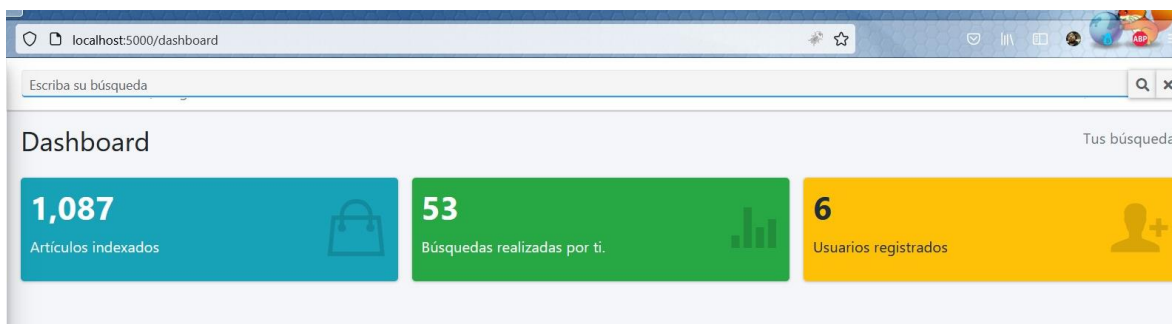
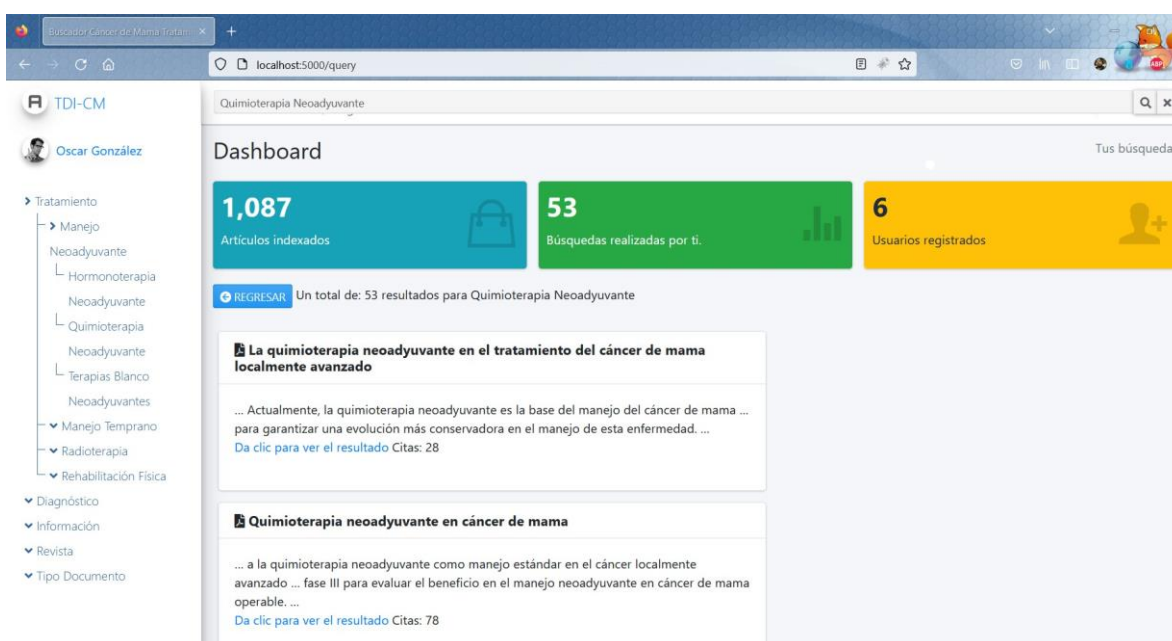


Figura 20 - Buscador

## 6.1.4 Resultados de una Búsqueda

Una vez que el usuario ha realizado una búsqueda, el sistema le devolverá aquellos recursos relacionados al concepto introducido en el buscador, en la parte inferior aparecerán tarjetas blancas con información relevante como el título, un breve resumen, # de citas, y un enlace para descargar el pdf o redirigir al usuario a un sitio externo para que obtenga el recurso, dado que puede ser de paga, aclaremos que los recursos descargables son meramente gratuitos, por lo que no se presenta ningún problema con alguna violación de información o derechos de autor, véase la Figura 21.



The screenshot shows a web browser window with the URL localhost:5000/query. The application interface includes a sidebar menu on the left with categories like 'Tratamiento', 'Diagnóstico', 'Información', 'Revista', and 'Tipo Documento'. The main content area displays a 'Dashboard' with three key metrics: 1,087 indexed articles, 53 searches performed by the user, and 6 registered users. Below the dashboard, search results are shown for 'Quimioterapia Neoadyuvante', including a summary of 28 citations and a summary of 78 citations.

Metric	Value
Artículos indexados	1,087
Búsquedas realizadas por ti.	53
Usuarios registrados	6

**REGRESAR** Un total de: 53 resultados para Quimioterapia Neoadyuvante

**La quimioterapia neoadyuvante en el tratamiento del cáncer de mama localmente avanzado**  
... Actualmente, la quimioterapia neoadyuvante es la base del manejo del cáncer de mama ... para garantizar una evolución más conservadora en el manejo de esta enfermedad. ...  
[Da clic para ver el resultado](#) Citas: 28

**Quimioterapia neoadyuvante en cáncer de mama**  
... a la quimioterapia neoadyuvante como manejo estándar en el cáncer localmente avanzado ... fase III para evaluar el beneficio en el manejo neoadyuvante en cáncer de mama operable. ...  
[Da clic para ver el resultado](#) Citas: 78

Figura 21 - Resultados de una búsqueda

## 6.1.5 Cerrar sesión

Si el usuario ha terminado de realizar sus búsquedas podrá cerrar su sesión por medio del botón habilitado en la parte superior derecha del Dashboard como se indica en la Figura 22.

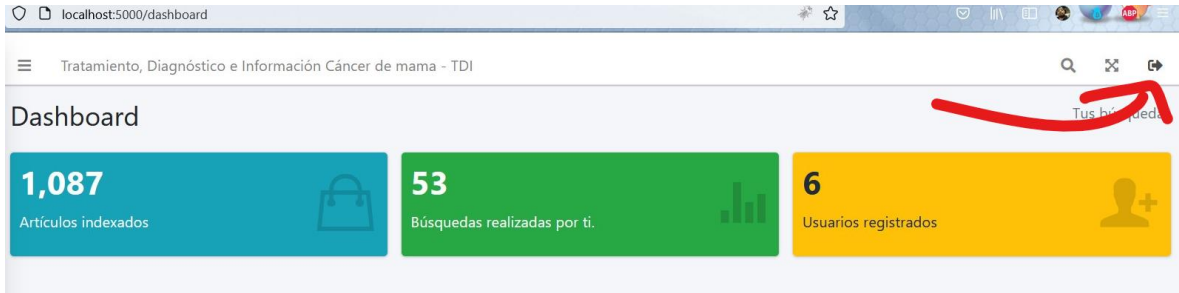


Figura 22 - Cerrar sesión

En la Figura 23, se puede apreciar, la manera en la que se almacena la información obtenida por el API, como se explica en la Tabla 2 se almacena el Título del recurso, su resumen, las palabras claves, el enlace de descarga o de acceso al recurso privado, y el tipo de documento con el fin de identificar qué tipo de recurso es.

idRecursos	Título	Resumen	KeyWords	Link	TipoDocumento
1	La quimioterapia neoadyuvante en el tratamiento de...	... Actualmente, la quimioterapia neoadyuvante es la...	Quimioterapia, Quimioterapia Neoadyuvante, Tratami...	<a href="https://www.scielo.org/article/ssm/content/raw/?...">https://www.scielo.org/article/ssm/content/raw/?...</a>	PDF
2	Quimioterapia neoadyuvante en cáncer de mama	... a la quimioterapia neoadyuvante como manejo está...	Cáncer de mama, Quimioterapia, Quimioterapia Neoad...	<a href="https://www.medigraphic.com/cgi-bin/new/resumen.cg...">https://www.medigraphic.com/cgi-bin/new/resumen.cg...</a>	PDF
3	Experiencia en el manejo del cáncer de mama en el ...	... a cirugía, por cáncer de mama, durante el perio.....	Cáncer de mama, Tratamiento, Cirugia	<a href="https://www.gamo-smeo.com/temp/GAMO%20V5%20No%205%...">https://www.gamo-smeo.com/temp/GAMO%20V5%20No%205%...</a>	WEB
4	Manejo de la axila tras el tratamiento neoadyuvant...	El cáncer de mama es un problema socio-sanitario d...	Cáncer de mama, Tratamiento Neoadyuvante, Quimio...	<a href="https://zaguan.unizar.es/record/111606">https://zaguan.unizar.es/record/111606</a>	PDF
5	La estadificación con resonancia magnética puede c...	... , la administración de quimioterapia neoadyuvant...	Cáncer de mama, Tratamiento, Quimioterapia, Tratam...	<a href="https://www.sciencedirect.com/science/article/pii/...">https://www.sciencedirect.com/science/article/pii/...</a>	PDF
6	Tratamiento neoadyuvante del cáncer de mama localm...	... a situaciones graves que pusieran en peligro la ...	Tratamiento, Cáncer de mama, Quimioterapia, Quimio...	<a href="https://www.redalyc.org/pdf/662/66276104.pdf">https://www.redalyc.org/pdf/662/66276104.pdf</a>	WEB
7	Manejo sistémico adyuvante en pacientes con cáncer...	... quimioterapia neoadyuvante, seguida de resección...	Cáncer de mama, Tratamiento, Quimioterapia, Quimio...	<a href="https://www.revistacancercol.org/index.php/cancer/...">https://www.revistacancercol.org/index.php/cancer/...</a>	WEB

Figura 23 - Almacén de información de los recursos

## Capítulo 7

### 7.1 Codificación

En este apartado se presenta una breve explicación de cómo es que se obtienen los recursos por medio del API y su integración con el desarrollo del Sistema de Recomendación.

#### 7.1.1 Estructura del Proyecto

En la Figura 24, se tiene la estructura del prototipo del proyecto, la cual se visualiza a modo de árbol jerárquico, teniendo en la carpeta **static** todos aquellos documentos y carpetas que servirán para darle funcionamiento y atractivo visual a la plataforma, aquí se encuentran almacenados todos los archivos JS, CSS3, JSON y SCSS. Dentro de la carpeta **templates** se tienen aquellos archivos HTML que servirán para dar estructura al sitio, los cuales se apropian de las características que se indiquen en los archivos de configuración almacenados en la carpeta *static*. Mientras que los archivos con extensión .py, son aquellos en los que se encuentra toda la parte lógica programable y que hace que el sistema obtenga los resultados conforme se realicen consultas, claro que, devolverá con prioridad aquellos que sean buscados si es que están en la taxonomía, en caso de que no se encuentre el concepto en la taxonomía, buscará pero no almacenará la información, solo la palabra de búsqueda, para que en el trabajo a futuro, el cual se describe más adelante, se puedan utilizar esos conceptos nuevos para insertarlos en la taxonomía. El archivo app.py contiene los métodos del controlador para devolver la información al usuario. El archivo mysql.py contiene los modelos y la conexión que se establece entre Python y mysql para que se puedan procesar las consultas. Mientras que en el archivo seartest.py se tiene el script que almacena el funcionamiento y las credenciales de uso del API para obtener los resultados de manera periódica.

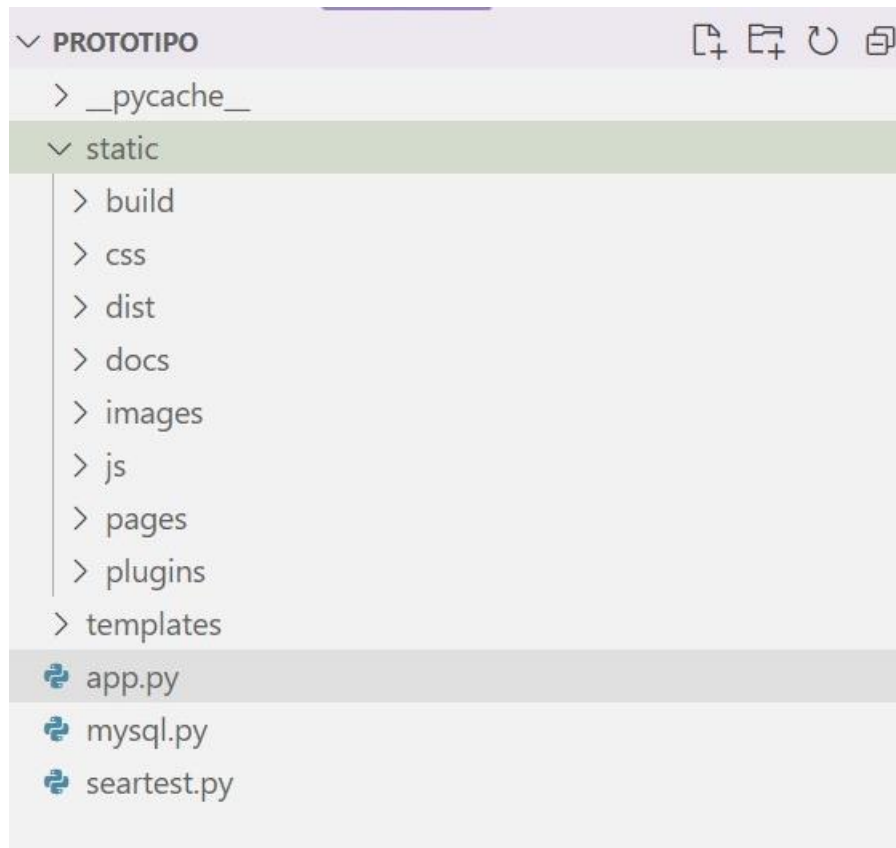


Figura 24 - Arquitectura del Proyecto

Flask, maneja un sistema de rutas para devolver la información y estructurar el sitio de manera eficiente por medio de palabras cortas o palabras clave de acceso a cada página que se tienen en el sistema, como se puede apreciar en la Figura 25, dentro del archivo `app.py` se declara en un inicio las librerías de Python a utilizar, así como se define el método principal de la aplicación y la ruta principal, siendo `'/'` la ruta del login para que un usuario inicie sesión. Se puede apreciar que en la instrucción `return render_template('login.html')` se indica a la ruta que tiene que devolver al usuario el archivo `login.html` el cual contiene la estructura en formato HTML5 del inicio de sesión.



```
pp.py > Index
from flask import Flask, render_template
from flask import Flask, render_template, request, redirect, url_for, flash
import pymysql

app = Flask(__name__)

# agregando rutas

@app.route('/')
def Login():
    return render_template('login.html')
```

Figura 25 - Librerías y Ruta Principal

En la Figura 26, se visualiza parte del contenido del archivo login.html, el cual contiene la estructura base en formato HTML5, la cual funciona como maqueta del sitio, para que junto con los archivos JS y CCS3 se le de forma y cuerpo al inicio de sesión.

```
templates > <> login.html > html > body.hold-transition.login-page > div.login-box > div.card.card-outline.card-primary > div.card-body > form
1 <!DOCTYPE html>
2 <html Lang="en">
3 <head>
4 <meta charset="utf-8">
5 <meta name="viewport" content="width=device-width, initial-scale=1">
6 <title>TDI-CM | Iniciar sesión</title>
7
8 <!-- Google Font: Source Sans Pro -->
9 <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Source+Sans+Pro:300,400,400i,700&display=fallb
10 <!-- Font Awesome -->
11 <link rel="stylesheet" href="{{url_for('static', filename='plugins/fontawesome-free/css/all.min.css')}}">
12 <!-- icheck bootstrap -->
13 <link rel="stylesheet" href="{{url_for('static', filename='plugins/icheck-bootstrap/icheck-bootstrap.min.css')}}">
14 <!-- Theme style -->
15 <link rel="stylesheet" href="{{url_for('static', filename='dist/css/adminlte.min.css')}}">
16 </head>
17 <body class="hold-transition login-page">
18 <div class="login-box">
19 <!-- /.Login-Logo -->
20 <div class="card card-outline card-primary">
21 <div class="card-header text-center">
22 <a href="../../index2.html" class="h1"><b>TDI</b><b>CM</b></a>
23 </div>
24 <div class="card-body">
25 <p class="login-box-msg">Inicio de sesión</p>
```

Figura 26 - Login.html



En la Figura 27, se aprecia la definición de la ruta para el Dashboard o tablero principal ('/dashboard'), el cual podrá recibir información por medio del método GET o POST, esta información es enviada desde un formulario HTML, se define un método llamado Index(), el cual consulta a la base de datos las facetas y todos aquellos términos relacionados a estas, para construir así la taxonomía que se muestra en la parte izquierda del dashboard. Con esto se construye la Taxonomía para que pueda ser visualizada. Devolviendo al final una vista, llamado al archivo 'index.html' pasándole como parámetros los resultados de las consultas para construir la taxonomía, así como algunos datos extra.

```

app.py > Index
14 @app.route('/dashboard', methods=['POST', 'GET'])
15 def Index():
16     connection = pymysql.connect(
17         | host="localhost", user="root", passwd="", database="modeloprevcanclit")
18     cursor = connection.cursor()
19     # queries for retrievint all rows
20     qry_facteta = "Select * from Faceta;"
21     # executing the quires
22     cursor.execute(qry_facteta)
23     faceta_rows = cursor.fetchall()
24     # Consulta solo Los Terminos con IDTermino NULL Padres
25     qry_terminos = "SELECT termino.idTermino, faceta.idFaceta, faceta.Nombre_Faceta, termino.Nombre_Termino, termino."
26     cursor.execute(qry_terminos)
27     term_rows = cursor.fetchall()
28     # Consulta solo Los Terminos con IDTermino != NULL Hijos
29     qry_terminosNN = "SELECT termino.idTermino, faceta.idFaceta, faceta.Nombre_Faceta, termino.Nombre_Termino, termino."
30     cursor.execute(qry_terminosNN)
31     term_rowsNN = cursor.fetchall()
32     #Consulta cuantos documentos hay
33     qry_terminosCount = "SELECT COUNT(Titulo) FROM recursos;"
34     cursor.execute(qry_terminosCount)
35     term_rowscOUNR = cursor.fetchall()
36     # committing the connection then closing it.
37     connection.commit()
38     connection.close()
39     return render_template('index.html', facetas=faceta_rows, terminos=term_rows, terminosh=term_rowsNN, countRec=ter
..

```

Figura 27 - Ruta Dashboard

En la Figura 28, se puede ver la construcción de la taxonomía a nivel de código que se encuentra en el archivo index.html, utilizando como parámetro los datos que se enviaron desde la ruta Dashboard, por medio de Jinja que es un motor de plantillas web para el lenguaje de programación Python, el cual facilita el control y acceso a los datos que se consultan desde el modelo y que se trabajan en el controlador para posteriormente en la vista mostrar al usuario los resultados.

```

<nav class="mt-2">
  <ul id="tree3">
    {% for faceta in facetas %}
    <li><a href="#">{{ faceta[1] }}</a>
      <ul>
        {% for termino in terminos %}
        {% if(termino[2]==faceta[1])%}
        <li><a href="#">{{ termino[3] }}</a>
          {% for terminoh in terminosh %}

          {% if(terminoh[4]==termino[0])%}
          <ul>
            <li><a href="#">{{ terminoh[3] }}</a>
              {% for terminoh1 in terminosh %}
              {% if(terminoh1[4]==terminoh[0])%}

```

Figura 28 - Código de la Taxonomía

En la Figura 29, se puede apreciar la ruta llamada *query* que consultará y devolverá los resultados al usuario, una vez que por medio de la interfaz haya introducido el concepto de la búsqueda, recalquemos que las búsquedas se realizan acorde a los conceptos que se tienen en la taxonomía. La primer consulta busca en la tabla de términos, aquellos que coincidan con la palabra o frase que ingreso el usuario, para posteriormente, en la segunda consulta, buscar aquellos documentos que estén relacionados con los términos que se recuperaron anteriormente, los resultados se procesan de acuerdo al contenido, destaquemos que hasta este apartado devolverá solamente aquellos recursos que ya existan en la base de datos previamente registrados por la búsqueda de cualquier otro usuario o que ya existieran desde el momento en que se realizaban las pruebas antes de integrar por completo el sistema. Más adelante se explica cómo se recuperan recursos nuevos en caso de no estar indexados para uno o varios términos de la taxonomía.

```

@app.route('/query', methods=['POST', 'GET'])
def Resultados():
    if request.method == 'POST':
        op = request.form['busqueda']
        #busqueda = op.split()
        busqueda = op
        print(busqueda)
        listTerm=list()
        listRec=list()
        # Consulta por buscador
        searchqry = "SELECT idTermino, Nombre_Termino, Descripcion_Termino FROM termino WHERE Nombre_Termino LIKE %s"
        cursor.execute(searchqry, f"%{busqueda}%")
        search_Qry = cursor.fetchall()
        for search_Qry1 in search_Qry:
            searchqryTer = "SELECT DISTINCT * FROM termino_has_recursos WHERE Termino_idTermino = %s"
            cursor.execute(searchqryTer, search_Qry1[0])
            search_QryTerlis = cursor.fetchall()
            for sear in search_QryTerlis:
                #print(sear)
                listTerm.append(sear[1])
            convert_list_to_set = set(listTerm)
        for itema in convert_list_to_set:
            return render_template('result.html', facetas=faceta_rows, terminos=term_rows, terminosh=term_rowsNN, query=busqu

```

Figura 29 - Consultas a la Base de Datos

Para recuperar documentos que no se hayan indizado aún en la taxonomía, se puede observar en la Figura 30, el procedimiento, en un inicio se declaran las librerías de Python a utilizar y las del API SerpApi, posteriormente se hace la conexión a la base de datos y se hace una consulta para traer todos los términos de la taxonomía ya que la palabra que ingrese el usuario si es similar o igual a la de un concepto de la taxonomía, los recursos recuperados los indizara a ese término. Mientras que en *params* se declaran las configuraciones o los parámetros que necesita el API que son, *engine*: el lugar o sitio de donde se extraerá la información, *q*: la query o búsqueda que se pretende encontrar en el sitio a scrapear, *api\_key*: la llave privada que se utiliza para poder utilizar el API. Se deben generar api-key cada determinado periodo de tiempo para poder seguir utilizando el API.

```

from serpapi import GoogleSearch
import pprint
import json
import mysql.connector
mydb = mysql.connector.connect(
    host="localhost",
    user="root",
    password="",
    database="modeloprevcanclit"
)
mycursor = mydb.cursor()
# Consulta solo Los Terminos con IDTermino
mycursor.execute(
    "SELECT termino.idTermino, termino.Nombre_Termino FROM termino WHERE 1")
term_rows = mycursor.fetchall()
# print(term_rows)
# for row in term_rows:
#     print(row[3])
# committing the connection then closing it.

params = {
    "engine": "google_scholar",
    "q": "COVID-19 cancer de mama",
    "api_key": "af2a74e994a44914dc3741d5e9dd187b8b8dc4d0b1a8f06679b68d0d9a4a"
}

```

Figura 30 - Recuperación de Recursos Parte 1

Para la obtención de recursos de manera periódica o incluso al momento de realizar una búsqueda por un usuario desde el sistema, utilizando el API, se llama a la función GoogleSearch con los parámetros descritos en el párrafo anterior, esos datos se pasan a una variable llamada results, que contendrá un diccionario de datos con todos aquellos resultados orgánicos de la búsqueda, los resultados orgánicos son aquellos que no tienen influencia de pago o algún otro tipo, son datos sin promoción, se transforman en formato JSON para ser manipulados con facilidad, para cada uno de los resultados, se almacena la información en la base de datos, posteriormente compara las cadenas de el titulo y la búsqueda que realizo el usuario con fin de saber si el resultado obtenido podría indizarse a algún otro concepto de la taxonomía. De esta manera se indizan los recurso con uno o más conceptos que tengan relación con la taxonomía, véase el código en la Figura 31.

```

search = GoogleSearch(params)
results = search.get_dict()
organic_results = results['organic_results']
y = json.dumps(organic_results)
# print(y)
prepo = ["de", "la", "con", "el", "en", "o", "a", "e", "a",
         "ante", "bajo", "cabe", "con", "contra", "desde"]
print('-----')
for res in organic_results:
    titulo = res['title'].lower()
    sqlinsert = """INSERT INTO recursos (Titulo, Resumen, link) VALUES (%s, %s, %s)"""
    mycursor.execute(sqlinsert, (res['title'], res['snippet'], res['link']))
    lastid=mycursor.lastrowid
    mydb.commit()
    for ter in term_rows:
        tersep = ter[1].lower().split()
        for teste in tersep:
            if teste not in prepo:
                resul = teste in titulo
                if resul:
                    print(teste+" esta en: "+titulo)
                    sqlinsert2 = """INSERT INTO termino_has_recursos(Termino_idTermino, Recursos_idRecursos) VALUES (%s,%s)"""
                    mycursor.execute(sqlinsert2, (int(ter[0]),lastid))

```

Figura 31 - Recuperación de Recursos Parte 2

Finalmente, por parte del código se manda a llamar a el método principal *main* para ejecutar el programa y poder visualizarlo en un navegador, se pasan como parámetros el puerto y el tipo de publicación como se puede apreciar en la Figura 32.

```

if __name__ == '__main__':
    app.run(port=3000, debug=True)

```

32 - Definición del método principal

## Capítulo 8

### 8.1 Resultados

Este trabajo implicó una búsqueda exhaustiva de información referente a los Sistemas de Recomendación, las taxonomías y la recuperación de información, al tratar toda esta información en conjunto se cuenta actualmente con una taxonomía sólida que trata sobre el Diagnóstico, Información y Tratamiento del Cáncer de mama, la cual puede ser consultada y de ayuda para expertos en el área, e incluso la pueden utilizar personas que no sean parte de rama de la medicina o que no tengan conceptos claros de lo que conlleva este importante tema, ya que, se tiene de igual manera un diccionario de conceptos el cual puede ser consultado por si se quiere saber el significado de alguno de los conceptos. La taxonomía funge como base para el Sistema de Recomendación, ya que todos los recursos que se obtengan estarán indizados de acuerdo con la similitud con el que se tiene de la búsqueda del usuario y los conceptos de la taxonomía.

Con el paso del tiempo y conforme se vaya utilizando la taxonomía, el corpus tendrá una tendencia de crecimiento, ya que gracias a los scripts de obtención de recursos se garantiza que se sigan buscando recursos de manera periódica.

Finalmente, se desarrolló un prototipo de Sistema de Recomendación de Literatura Académica para artículos científicos médicos en el ámbito de cáncer de mama, específicamente diagnóstico y tratamiento, teniendo como base la taxonomía implementada a partir del Consenso Mexicano del Cáncer de mama. Este prototipo de Sistema de Recomendación pretende ser de apoyo para cualquier persona que desee obtener información relevante, puntual y precisa acerca del Cáncer de mama.

## Capítulo 9

### 9.1 Conclusiones y Trabajo a futuro

Existen diversos Sistemas de Recomendación no todos del mismo tipo, y mucho menos relacionados a los mismos temas, sin embargo, todos con un objetivo similar, proveer información puntual y precisa a los usuarios que la necesiten. Este trabajo, a pesar de ser un prototipo cuenta ya con un poderoso motor de recuperación de recursos digitales referentes al cáncer de mama, al igual que se cuenta con una amplia y basta taxonomía, la cual puede ser utilizada para otros proyectos relacionados.

Al ser un sistema de recomendación desarrollado bajo un marco web, la administración de este resulta ser muy sencillo, además de que al tener como base el lenguaje de programación Python haría que el código existente sea fácil de escalar por cualquier programador que conozca el lenguaje.

Se ha aprendido bastante acerca de lo que son las taxonomías y como utilizarlas al igual que con los Sistemas de Recomendación, al existir distintos tipos, cualquiera podría ser apto para soportar información como la que se pretende obtener y otorgar a los usuarios, claro que con el apoyo de una taxonomía para poder lograr la relación entre términos, búsquedas y recursos.

Con todo lo implementado y desarrollado, se tiene un prototipo funcional, este proyecto puede tener un crecimiento evolutivo acorde con las necesidades y funcionamiento de los Sistemas de Recomendación, teniendo, como trabajo a futuro, actualizaciones en el Sistema para mejorar el funcionamiento, tanto para la recomendación de recursos dada una búsqueda, así como para el análisis de sus recursos y así tener una indización más precisa.

Actualizar el Sistema de Recuperación para tener usuarios de tipo administrador que sean expertos en el área de cáncer de mama y puedan determinar si nuevos conceptos pueden ser agregados a la taxonomía e incluso poder modificarla desde el mismo sistema, también contar con un administrador de usuarios para que, a partir de esto, el Sistema de Recomendación pueda evolucionar y tener un funcionamiento más completo.



## Bibliografía

- Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., & Xia, F. (2019). *Scientific Paper Recommendation: A Survey* (Vol. 7). IEEE Access. doi:10.1109/ACCESS.2018.2890388
- Bansal, S., & Baliyan, N. (2019). *A Study of Recent Recommender System Techniques* (Vol. 10). International Journal of Knowledge and Systems Science . doi:http://doi.org/10.4018/IJKSS.2019040102
- Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). *Research-paper recommender systems: a literature survey* (Vol. 17). International Journal on Digital Libraries volume. doi:https://doi.org/10.1007/s00799-015-0156-0
- Beel, J., Langer, S., Gipp, B., & Nürnberger, A. (2014). *The Architecture and Datasets of Docear's Research Paper Recommender System* (Vol. 20). D-Lib Magazine. doi:https://dx.doi.org/10.1045/november14-beel
- Cárdenas S., J. (2021). *Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario*. Colima: Gaceta Mexicana de Oncología.
- Galán Nieto, S. M. (2007). Filtrado Colaborativo y Sistemas de Recomendación. 8.
- Gauchat, J. (2012). *El gran libro de HTML5, CSS3 y Javascript*. Barcelona, España: Marcombo. Retrieved from El gran libro de HTML5, CSS3 y Javascript.
- González Escamilla, O., & Marcellin Jacques, S. (2017). *Estado del arte en los sistemas de recomendación* (Vol. 135). Res. Comput. Sci.
- H. Canós, J., Letelier, P., & Penadés, M. (2007). Metodologías Ágiles en el Desarrollo de Software. *DSIC*, 9.
- Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). *Recommending And Evaluating Choices In A Virtual Community Of Use*. Morristown: Proceedings of the SIGCHI conference on Human factors in computing systems. Retrieved from https://dl.acm.org/doi/fullHtml/10.1145/223904.223929



- Jallouli, M., Lajmi, S., & Amous, I. (2017). *Designing Recommender System: Conceptual Framework and Practical Implementation* (Vol. 112). Procedia Computer Science. doi:<https://doi.org/10.1016/j.procs.2017.08.195>
- Jugovac, M., Jannach, D., & Lerche, L. (2017). *Efficient optimization of multiple recommendation quality factors according to individual user tendencies*. (Vol. 81). Expert Systems with Applications. doi:<https://doi.org/10.1016/j.eswa.2017.03.055>
- LLC, S. (2022). *SerpApi*. Retrieved from SerpApi: <https://serpapi.com/>
- Manouselis, N., & Verbert, K. (2013). *Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation* (Vol. 18). Procedia Computer Science.
- Martín, E. V. (2016). *Sistemas de recomendación semánticos para la compartición de conocimiento y la explotación de tesauros. Un enfoque práctico en el ámbito de los sistemas nutricionales*. España: Universidad de Granada.
- Mendoza Olguín, G. E., Laureano de Jesús, Y., & Pérez de Celis Herrero, M. (2019). *Métricas de similaridad y evaluación para sistemas de recomendación de filtrado colaborativo* (Vol. 7). Puebla, Puebla, México: Revista de Investigación en Tecnologías de la Información: RITI.
- Muthukadan, B. (2011-2018). *Selenium with Python*. Retrieved from Selenium with Python: <https://selenium-python.readthedocs.io/>
- Nilashi, M., Ibrahim, O., & Bagherifard, K. (2018). *A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques* (Vol. 92). Expert Systems with Applications. doi:<https://doi.org/10.1016/j.eswa.2017.09.058>
- Núñez V., E. R. (2012). *Sistemas de recomendación de contenidos para libros inteligentes*. Universidad de Oviedo.
- O'Mahony, M., Hurley, N., Kushmerick, N., & Silvestre, G. (2004). *Collaborative recommendation: A robustness analysis* (Vol. 4). Internet Technol. doi:<https://doi.org/10.1145/1031114.1031116>

- Pallets. (2010). *Flask*. Retrieved from Flask: <https://flask.palletsprojects.com/en/2.2.x/>
- Park, D., Kim, H., Choi, I., & Kim, J. (2012). *A literature review and classification of recommender systems research* (Vol. 39). Expert Systems with Applications. doi:<https://doi.org/10.1016/j.eswa.2012.02.038>
- Peis, E., Morales del Castillo, J. M., & Delgado López, J. A. (2008). Sistemas de Recomendación Semánticos. Un análisis del estado de la cuestión. *Hipertext*, 18.
- Pera, M. S., & Ng, Y. K. (2014). *Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles* (Vol. 42). Journal of Intelligent Information Systems volume . doi:<https://doi.org/10.1007/s10844-013-0298-8>
- Pérez Almaraz, A., & Close G., J. (2015). *Sistemas de recomendación. Robots Autónomos: Navegación*.
- Python Software Foundation. (2001-2022). *Python*. Retrieved from Python: <https://www.python.org/>
- Rich, E. (1979). *User modeling via stereotypes* (Vol. 3). Cognitive Science. doi:[https://doi.org/10.1016/S0364-0213\(79\)80012-9](https://doi.org/10.1016/S0364-0213(79)80012-9)
- Richardson, L. (2004-2015). *Beautiful Soup*. Retrieved from Beautiful Soup: <https://beautiful-soup-4.readthedocs.io>
- Rici, F., Rokach, L., & Shapira , B. (2010). *Introduction to Recommender Systems Handbook*. Boston: Springer. doi:[https://doi.org/10.1007/978-0-387-85820-3\\_1](https://doi.org/10.1007/978-0-387-85820-3_1)
- Seguido Font, M. (2009). *Sistemas de recomendación para webs de información sobre la salud*. Catalunya: Departamento de Lenguajes y Sistemas Informáticos.
- Silberschatz, A., Korth, H., S. Sudarshan, & Sudarshan, S. (2002). *Fundamentos de bases de datos*. Madrid: McGraw-Hill. Retrieved from <https://www.urbe.edu/UDWLibrary/InfoBook.do?id=9188>

- Stanford University. (2016-2020). *Protégé*. Retrieved from Protégé:  
<https://protege.stanford.edu/>
- Sugiyama, K., & Kan, M. Y. (2010). *Scholarly Paper Recommendation via User's Recent Research Interests*. New York: Association for Computing Machinery. doi:10.1145/1816123.1816129
- Sun, J., Ma, J., Liu, Z., & Miao, Y. (2014). *Leveraging Content and Connections for Scientific Article Recommendation in Social Computing Contexts* (Vol. 57). *The Computer Journal*. doi:10.1093/comjnl/bxt086
- Trappey, A. J., Trappey, C. V., Wu, C. Y., Fan, C. Y., & Lin, Y. L. (2013). *Intelligent patent recommendation system for innovative design collaboration*, (Vol. 36). *Journal of Network and Computer Applications*. doi:<https://doi.org/10.1016/j.jnca.2013.02.035>
- Wang, D., Liang, Y., Xu, D., Feng, X., & Guan, R. (2018). *A content-based recommender system for computer science publications* (Vol. 157). *Knowledge-Based Systems*. doi:<https://doi.org/10.1016/j.knosys.2018.05.001>
- Xia, F., Liu, H., Lee, I., & Cao, L. (2016). *Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences* (Vol. 2). *IEEE Transactions on Big Data*. doi:10.1109/TBDATA.2016.2555318