

Sistematización sobre la generación de perfiles de preferencias de usuarios

Yaili Ledea-Velázquez^{1*}, Hubert Viltres Sala²

¹Universidad de las Ciencias Informáticas; La Habana, Cuba. yledea@uci.cu

²Universidad de las Ciencias Informáticas, La Habana, Cuba. hviltres@uci.cu

Recibido: 25 agosto 2022

Aceptado: 11 oct. 2022

RESUMEN

Los sistemas de recuperación de información facilitan el acceso a la información disponible en más de un billón de sitios web en Internet. El crecimiento exponencial de la información y la diversidad de preferencias de los usuarios plantean la necesidad de aplicar técnicas que permitan comprender la intención de búsqueda y brindar información relevante y personalizada. Este trabajo tiene como objetivo determinar las tendencias actuales en la identificación y generación de perfiles de usuarios. Con este fin se realizó una revisión de literatura organizada en tres etapas. La literatura analizada comprende artículos de revistas, conferencias, tesis y libros en idioma inglés y español desde el 2016 hasta agosto del 2022. Se obtuvo como resultado una sistematización de criterios de diferentes autores y estudios de casos satisfactorios que reveló los principales referentes teóricos que sustentan la generación de perfiles de preferencias de usuarios.

PALABRAS CLAVE: Perfiles de preferencias de usuarios; Recuperación de información; Revisión sistemática

ABSTRACT

Systematization on the generation of user preference profiles. There are more than a billion websites and this number continues to grow exponentially. It is common to use retrieval systems to find information that meets the needs of users. To improve the relevance of the results obtained, techniques such as the generation of user preference profiles are used, among others. The objective of this work is to determine the current trends in the implementation and development of user profiles. To this end, a literature review organized in three stages was carried out. The literature analyzed includes articles from scientific journals, conferences, theses and books in English and Spanish from 2016 to date. The result was a systematization of criteria from different authors and satisfactory case studies that revealed the main theoretical references that support the generation of user preference profiles.

KEYWORDS: User preference profiles; Information retrieval; Systematic review

INTRODUCCIÓN

El surgimiento de nuevas tecnologías para mejorar el acceso a la información y el crecimiento exponencial de Internet aumentan el volumen de contenidos web disponible. Datos correspondientes a 2019 reflejan el constante aumento del flujo de información en la red; donde, en sólo 60 segundos se comparten ahora, más de 87 mil 500 tuits en Twitter, 4.8 millones de videos en YouTube y se transfieren alrededor de 900 mil dólares por compras online (*Lo que ocurre en Internet en un minuto*, 2019). La información en la web se encuentra dispersa y poco estructurada o es invisible al usuario común dificultando el acceso a la información que más se ajusta a sus necesidades por lo que es común el empleo de Sistemas de Recuperación de Información (SRI) para encontrar lo buscado.

Existe coincidencia respecto a que el objetivo principal de un SRI es satisfacer la necesidad de información planteada por un usuario en una consulta en lenguaje natural especificada a través de un conjunto de palabras clave, que ayudan a identificar la información más relevante para el usuario (Zhai & Lafferty, 2017; Sala et al., 2019). Cada usuario puede tener diferentes necesidades y objetivos bajo una misma consulta que en la mayoría de las ocasiones es corta y ambigua (Yee, 2020). Para mejorar la relevancia de la información obtenida por los SRI se reporta el uso de varias técnicas, relacionadas con la expansión y desambiguación de consultas, el uso de bases de conocimiento (ej. taxonomías, ontologías, etc.) y la generación de perfiles de preferencias de usuarios (Hernández & Sala, 2018).

La palabra perfil según la Real Academia Española tiene varias acepciones, entre ellas “contorno de la figura de algo o de alguien; conjunto de rasgos peculiares que caracterizan a alguien o algo; miramientos en la conducta o en el trato social” (RAE, 2022). En el mundo de la tecnología los perfiles describen las características más vitales del objeto en forma compacta, de manera que los ordenadores lo entiendan para que los programas los

puedan usar en sus tareas. En consecuencia, el perfilado o *profiling* es una técnica que tiene como idea principal la de crear y gestionar perfiles de usuarios.

Kanoje et al. (2016) definen el perfilado de usuarios como el proceso de extraer, integrar e identificar información basada en palabras clave para generar un perfil estructurado y luego visualizar el conocimiento extraído de esos hallazgos. El perfilado de usuarios en la web es el proceso de obtención de valores de diferentes propiedades que constituyen el modelo de usuario. Un sistema típico de perfilado está orientado a la búsqueda extracción e integración del perfil basado en palabras clave de la web.

Existe consenso entre los trabajos de Rodríguez et al. (2016), Singh y Sharma (2018) y Viltres et al. (2018) al especificar que el perfil de usuario puede contener tanto datos personales y características del sistema computacional como patrones de comportamiento, intereses personales y preferencias. Este modelo de usuario está representado por una estructura de datos adecuada para su análisis, recuperación y utilización. Permite determinar el contexto y la intención del usuario para que el SRI seleccione los documentos más relevantes y dentro de la computadora representa la “forma física” de las necesidades de información del usuario.

Según los distintos trabajos consultados (EKE et al., 2019; Gao, 2019; Lashkari et al., 2019; Sala et al., 2019; Grida et al., 2020) y a los efectos de la presente investigación se define el perfil de preferencias de usuarios como un modelo que representa el comportamiento de los usuarios al acceder a un SRI; que, además, están en constante cambio y actualización, para garantizar que la comunicación de entre el usuario y el sistema garantice la obtención de información relevante y personalizada.

Los perfiles de preferencias de usuarios son utilizados por sistemas de recomendación en el filtrado de información (películas, música, libros, noticias, páginas web, etc.) para mostrar sugerencias relevantes, según su perfil de preferencias (gustos, intereses, necesidades), sin necesidad de realizar búsquedas explícitas de estos recursos (N. Rodríguez et al., 2017). Los Sistemas de Recuperación de Información (SRI) utilizan los perfiles de preferencias de usuarios para el procesamiento de consultas, para identificar su necesidad y personalizar los resultados de búsqueda (Viltres et al., 2018).

Los perfiles de usuario son creados a partir de sus intereses, preferencias y necesidades de información, y pueden ser empleados por los sistemas de recuperación para maximizar la satisfacción del usuario con los resultados obtenidos. El presente trabajo se centra en analizar las principales técnicas, métodos y tendencias existentes en la actualidad para la generación de perfiles de preferencias de usuario.

Materiales y Métodos

La presente investigación constituye una revisión sistemática de la literatura sobre los principales métodos, modelos y algoritmos empleados en la generación del perfil de preferencias de un usuario en la actualidad. Las revisiones sistemáticas de la literatura son estudios científicos de investigación secundaria e integrativa que brindan un resumen fiable, válido y actualizado de la mejor evidencia científica disponible respecto a determinado tema. Esta revisión se basa en el modelo propuesto por (Carrizo et al., 2018) que se define en tres bloques fundamentales en la investigación:

1. Definición para la búsqueda: donde se define la pregunta de investigación, el alcance de la revisión, los criterios de inclusión y exclusión, y finalmente la cadena de búsqueda.
2. Ejecución de la búsqueda: donde se define la selección de trabajos primarios y difusión de criterios de análisis.
3. Discusión de los resultados: donde se definen los esquemas de caracterización y se analizan los resultados.

Este trabajo tiene como objetivo fundamental determinar las tendencias de implementación y desarrollo en la generación de perfiles de preferencias de usuario, con el fin de mejorar la relevancia de los resultados de búsqueda en un Sistema de Recuperación de Información. Para la búsqueda información se empleó mayormente la base de datos *Google Scholar* (GS) pues es muy sensible y fácil de investigar, además de brindar un motor de búsqueda robusto y gratuito, geográficamente neutral y que abarca un volumen de información considerable. También se incluyeron materiales de *Elsevier*, *Research Gate* y repositorios institucionales.

La literatura analizada comprende artículos de revistas, conferencias, tesis, documentos y libros en idioma inglés y español. La selección de las fuentes bibliográficas se realizó mediante el filtrado selectivo de la información.

Inicialmente se seleccionaron las fuentes con título y resumen vinculado al tema de investigación y con fecha de publicación perteneciente al período 2016-2022. Luego se seleccionó solo la literatura que permitió su lectura de forma gratuita y completa. Este último filtrado garantizó la exclusión de aquellas fuentes bibliográficas que abarcan la temática desde perspectivas no relacionadas con el tema del presente estudio. Como resultado de este proceso de inclusión y exclusión finalmente se obtuvieron 60 fuentes bibliográficas, de las cuales 33 son artículos de revistas, 13 conferencias, 11 tesis, 1 documento y 2 libros.

RESULTADOS Y DISCUSIÓN

En esta sección se analizan y clasifican los resultados obtenidos a partir del estudio de las fuentes bibliográficas consultadas. Estos resultados fueron clasificados según la distribución de las publicaciones por año, los países y áreas geográficas donde se redactaron, las metodologías de investigación utilizadas y los temas investigados.

Los resultados obtenidos permiten determinar una serie de aspectos que constituyen el eje fundamental para realizar el análisis sobre perfiles de preferencias de usuarios en el período 2016-2022. La discusión se organiza en cuatro puntos fundamentales: análisis de la bibliografía consultada; contenido y comportamiento frente a los cambios, formas de adquisición de la información y modelado y construcción de los perfiles de preferencias de usuarios; limitaciones de la investigación y sugerencias para futuras investigaciones.

Distribución de las publicaciones por año

La Figura 1 muestra el período de tiempo al que pertenecen los artículos consultados, comprende desde enero de 2016 hasta agosto de 2022. El año 2018 es el de mayor cantidad de publicaciones y representa el 21.67% del total de los trabajos publicados durante todo el período, después de este año, las publicaciones sobre el tema analizado disminuyen. El 2020 es el año con menor número de trabajos publicados y durante el 2022 se identificó hasta el momento en que se realiza este estudio, cinco artículos. Las revistas científicas fueron los medios de publicación más empleados (55%), seguidas por las conferencias (21.67 %).

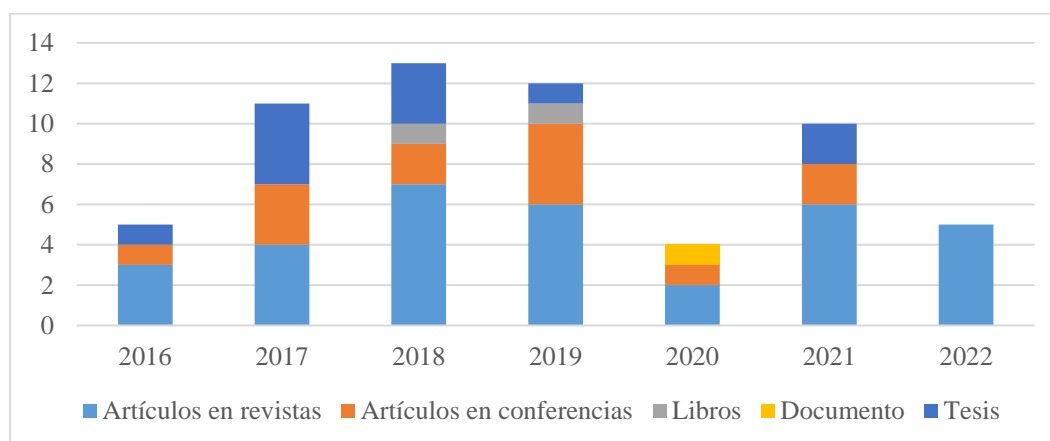


Figura 1. Distribución de las publicaciones por año. Elaboración propia

Distribución de las publicaciones por países y áreas geográficas

Los países líderes en la elaboración de estas publicaciones son India y China con el 13.3% y 10% respectivamente del total de las publicaciones. La Figura 2 muestra el nivel de publicaciones por cada área geográfica. El continente asiático es el líder en publicaciones (30%), seguido por América Latina y el Caribe (28.3%). En el estudio realizado se identificaron solo 2 publicaciones provenientes de Australia y Oceanía (3.3%).

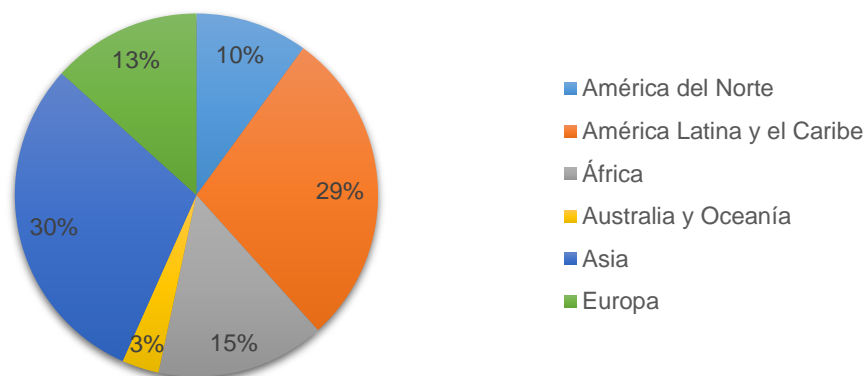


Figura 2. Distribución de las publicaciones por áreas geográficas. Elaboración propia

Distribución de publicaciones según métodos de investigación

El análisis de las publicaciones consultadas permitió determinar que los métodos de investigación más empleados son los estudios de casos (53.3%), seguidos por los estudios exploratorios (18.3%), las encuestas y las revisiones de la literatura (representando un 15% y 13.3% de los métodos respectivamente). Las revisiones de la literatura como los artículos que emplean las encuestas como método de investigación centran sus análisis en el modelado de los perfiles de preferencias de usuarios, así como la información que se almacena en un perfil y las formas de adquirirla. Las publicaciones basadas en estudios de casos, abordan en su mayoría la implementación de un modelo de perfil de usuario determinado y los resultados obtenidos.

Discusión de los resultados del análisis de la bibliografía consultada

El estudio realizado determinó que en la base de datos GS son los artículos publicados en revistas científicas los que más referencias hacen al tema de los perfiles de preferencias de usuarios en el período 2016-2022. Determinar con exactitud el país líder en publicaciones es un tema complejo ya que depende en gran medida del alcance

de la base de datos científica empleada. Según los estudios realizados por (Gao, 2019) son los países de habla inglesa los que más publicaciones ostentan seguidos por China. En contraste con estos autores en la presente investigación se determina que la India es el país con más publicaciones referentes a los perfiles de preferencias de usuarios en el periodo. Se considera que esta diferencia pudiera estar determinada por el empleo de diferentes bases de datos científicas y periodos de tiempo variables. Aunque existe una ligera discrepancia de criterios, se puede afirmar que el continente asiático destaca por la cantidad de publicaciones sobre el tema analizado.

Referente a los temas de investigación, el papel de los perfiles de usuarios en la personalización de los resultados de búsqueda y en sistemas de recomendación son los temas más investigados. Estos estudios hacen hincapié en la información almacenada en un perfil, así como la forma de recolectarla, los modelos existentes para el perfilado y su construcción.

Contenido y comportamiento frente a los cambios de los perfiles de preferencias de usuarios

Quiroz et al (2018) puntualizan que un perfil de usuario consta de cinco elementos:

- Datos personales (nombre, edad, dirección, género): es la información personal del usuario, esta información es explícitamente ofrecida por el usuario mediante cuestionarios.
- Estilo de aprendizaje (Activo/Reflexivo, Sensitivo/Intuitivo, Inductivo/Deductivo, Visual/Verbal, Secuencial/Global): se obtiene utilizando el test de Felder y Silverman en los usuarios.
- Historial de búsqueda.
- Interfaz de usuario: es la forma de interacción que posee el usuario con el sistema, esta puede ser modificada por el usuario o adaptada por el sistema en función a sus gustos y preferencias.

- Preferencias del usuario: se obtienen mediante cuestionarios realizados al usuario o implícitamente al analizar sus interacciones con el sistema.

La información almacenada en un perfil de usuario tiene diferente comportamiento frente a los cambios y se clasifican en estáticos y dinámicos (EKE et al., 2019). Los perfiles estáticos son aquellos que mantienen información del usuario que es muy poco probable que cambie en el tiempo y no está sujeta a continuas actualizaciones. En esta clasificación se agrupan los datos demográficos; conocimientos y habilidades de fondo y; necesidades y objetivos del usuario (Farid et al., 2018; Grida et al., 2020).

El mantenimiento de este tipo de información permite al sistema agrupar a los usuarios en estereotipos y tomar decisiones de personalización de alto nivel por ejemplo adaptar servicios basado en la localización geográfica del usuario (Niu et al., 2018). El problema con este tipo de perfiles es que los usuarios raramente proveen toda su información con precisión debido a la importancia que le confieren a su privacidad, lo que le resta confiabilidad.

El comportamiento del usuario, sus intereses y preferencias y su contexto son un ejemplo de la información que se recopila en los perfiles dinámicos. Los perfiles dinámicos son aquellos que realizan un seguimiento de la información que evoluciona con el tiempo (Vu, 2017).

El comportamiento es una información que se reúne implícitamente y su análisis permite la detección de patrones que pueden ser empleados por los sistemas inteligentes o adaptativos para determinar la estrategia de personalización a seguir. Algunos elementos a tener en cuenta para el estudio del comportamiento del usuario en el sistema son: tiempo de atención, elementos clicados y movimientos del mouse.

Los intereses y preferencias de los usuarios son una parte fundamental en la personalización y pueden estar definidos por temas sociales, sitios web, documentos o temas laborales. Estos elementos pueden derivarse de varias fuentes como el historial de compras o el historial de navegación.

Un usuario puede preferir unos enlaces o resultados de búsqueda en lugar de otros. Estas preferencias pueden ser persistentes o relativas, esto, depende de la necesidad actual del usuario, del objetivo y del contexto actual en general. Los SRI tienen en cuenta sus preferencias para devolver para mostrar al usuario respuestas específicas y acorde a sus intereses, sean estos, recientes o persistentes (Kacem, 2017).

Los modelos que mantienen intereses de usuario de corto plazo son usualmente creados sobre la marcha y son actualizados frecuentemente durante la sesión de búsqueda del usuario. Los intereses a largo plazo pueden ser considerados como información dinámica también si el sistema tiene un mecanismo de revisión o actualización para ellos (ej. aumento o disminución de ponderaciones de los intereses periódicamente) (Sinnamon et al., 2021). Más decisiones de personalización enfocadas en el usuario pueden realizarse cuando el sistema mantiene información dinámica que atiende el contexto actual e intereses del usuario.

Métodos de adquisición de la información

Varios autores (Costanzo et al., 2019; Gatziolis et al., 2022; Lekshmi & George, 2016; Liu et al., 2020) defienden que, para crear un perfil de un usuario con sus gustos, preferencias e intereses, se pueden utilizar dos formas o métodos en la recolección de información: implícita o explícita. En el acercamiento explícito el usuario crea su perfil a través de información que ingresa al sistema en respuesta a peticiones realizadas por este. Esta información puede contener una especificación inicial de sus temas de interés y de aquellos que no le interesan, retroalimentación positiva o negativa sobre la relevancia de documentos recuperados. El usuario también puede examinar (inspeccionar o modificar) la información que el sistema haya aprendido sobre él hasta el momento. Lo preocupante sobre este método es que el usuario puede no desear emplear tiempo extra en proveer información al sistema o que la información suministrada pueda poseer inconsistencias o errores.

eBay pide a los usuarios que proporcionen sus opiniones y valoraciones de los servicios y productos que se ofrecen, la empresa utiliza esta información para mejorar las recomendaciones que da a los clientes. Yahoo! solicita explícitamente al usuario que proporcione información personal que será almacenada para crear un perfil. Los perfiles explícitos tienen también una naturaleza estática y son válidos sólo hasta que el usuario cambie sus preferencias. Para superar este problema se sugiere extraer la información implícitamente mediante el seguimiento de las actividades y/o comportamiento de los usuarios.

En el método implícito la información es adquirida sin esfuerzo alguno por parte del usuario. A diferencia del perfilado estático, el perfilado dinámico utiliza el método implícito y analiza los patrones de comportamiento del usuario para determinar sus intereses y desarrollar conocimiento de fondo del usuario. Normalmente el sistema registra el historial de búsqueda del usuario en términos de las consultas introducidas y los resultados consultados, también incluye el procesamiento de cualquier documento o ítem almacenado (calendario, correo electrónico, etc.) o la recopilación de información de las interacciones del usuario con aplicaciones sociales (redes sociales, aplicaciones de etiquetado social, blogs, entre otros). Este método intenta inferir automáticamente los intereses o contexto del usuario a partir de los registros o ítems procesados por el usuario.

Es importante señalar que estos métodos no son mecanismos aislados, y muchas veces se utilizan simultáneamente (métodos híbridos), para producir perfiles más precisos y comprensibles. El enfoque híbrido recopila la información mediante técnicas semiautomatizadas con participación restringida del usuario, con un enfoque mixto de métodos implícitos y explícitos. Este enfoque ayuda a perfilar eficientemente y mantiene una precisión temporal del perfil a medida que la información se actualiza.

Métodos	Ventajas	Desventajas	Herramientas de recolección	Comportamiento frente al cambio	Información recolectada
Explícito	Incorpora información personal, retroalimentación	Tedioso para el usuario, información incierta o incompleta	Formulario de registro, cuestionarios, encuestas	Estático	Datos demográficos, conocimientos y habilidades, objetivos y necesidades
Implícito	No requiere esfuerzo por parte del usuario	Invasión a la privacidad, alta dependencia de interacción previa (<i>cold start problem</i>)	Registros de acceso y búsqueda, historial de navegación, vínculos visitados, etc.	Dinámico	Comportamiento, intereses y preferencias, otros contextos
Híbrido	Incorpora fortalezas de ambos métodos, reduce la necesidad de interacción previa al incorporar información personal		Técnicas semiautomatizadas con participación restringida del usuario	Dinámico	Ambos

Tabla 1: Resumen métodos de adquisición de la información y su comportamiento frente a los cambios

La Tabla 1 sintetiza las formas de recolección de información a través de sus ventajas y desventajas y las herramientas de recolección empleadas, así como su clasificación de acuerdo a cómo se manejan los cambios. Se puede considerar que el método híbrido es más eficiente y mejora la precisión de los resultados mostrados, pues parte de los datos introducidos inicialmente por el usuario mitigando en cierto grado el problema de inicio en frío a la vez que se actualiza continuamente para reflejar las preferencias temporales del usuario.

Modelado y construcción de perfiles de preferencias de usuarios

Existen diferentes clasificaciones de modelos de perfiles de usuario, entre ellos destacan los basados en palabras clave (Zhou et al., 2016), modelo de perfil semántico (Farid et al., 2018) y basados en conceptos o jerarquías (EKE et al., 2019).

Basados en términos o palabras clave (*Keyword-based*): La representación de perfiles de usuario por medio de palabras clave ponderadas o basado en vectores es una de las técnicas más empleadas. Estas palabras clave pueden ser extraídas automáticamente o proporcionadas por el usuario. Las ponderaciones o pesos son representaciones numéricas asociadas a las palabras clave, que representan intereses del usuario, las cuales son utilizadas para representar la importancia de ese tema de interés (Yee, 2020). La Frecuencia del Término (TF) y la Frecuencia Invertida del Término (TF-IDF) son métodos empleados para desarrollar una representación de cada término ponderado de los documentos o consultas (Bhavadarani et al., 2019). Además, la similitud entre dos vectores de términos ponderados es calculada por la Similitud del Coseno (CS) (Leyva et al., 2018). En este modelo la efectividad de los perfiles de usuario depende del grado de generalización del vector. Su principal desventaja es la polisemia que puede causar una pobre interpretación de los intereses del usuario.

Modelo de perfil semántico (*Semantic-Network-based or Ontology-based*): utiliza ontologías para representar la semántica en un perfil de usuario. Una ontología presenta un perfil como un conjunto de conceptos con las asociaciones entre estos conceptos dentro de un dominio. Las ontologías son consideradas como una opción superior para la próxima generación de perfiles de usuario debido a la gran representación del conocimiento y herramientas de inferencia relacionadas (Grida et al., 2020). Este modelo ha sido utilizado en diferentes dominios incluyendo motores de búsqueda personalizada (Abdeljaber, 2018; Tchantchou Samen & Ezin, 2019), recuperación de información (Rafa & Kechid, 2019), turismo (Garrido et al., 2017), *e-learning* (Adib et al., 2022; González, 2019) y comercio electrónico (Calvopiña & Pérez, 2018; García, 2021).

Basado en conceptos o jerarquías (*Concept-based or Hierarchy-based*): este tipo de perfiles funcionan igual que los perfiles basados en redes semánticas, en la representación por nodos de conceptos y relaciones entre esos nodos. Varios mecanismos son aplicados para asociar un peso al nivel de interés del usuario en cada tema. (Gu et al., 2020) proponen un marco de trabajo para resolver el problema de los perfiles de usuarios jerárquicos en sistemas de recomendación de comercio electrónico (HUP). Para su propuesta se auxilian de redes neuronales recurrentes piramidales, equipadas con *Behavior-LSTM* para formular los intereses jerárquicos en tiempo real de los usuarios en múltiples escalas. Además, en lugar de simplemente utilizar el comportamiento a nivel de elementos de los usuarios (ej.: calificaciones o clics) como en métodos más convencionales, HUP recolecta información secuencial de las interacciones temporales finamente granulares de los usuarios para modelar (micro comportamiento, ej.: clics en componentes de ítem como imágenes o comentarios, navegación en buscadores o recomendaciones).

Safi, Jaoua y Belguith (2017) proponen un método de ranking personalizado basado en un modelo de aprendizaje supervisado aplicado a la lengua árabe. Este método consta de cuatro pasos, el modelado del usuario, el pareo de documento-consulta-perfil, el aprendizaje para clasificar y la clasificación de resultados. Se formula un enfoque híbrido del modelado del usuario que se basa en representaciones conceptuales y multidimensionales mediante la explotación de los recursos semánticos árabes. En este enfoque, se introduce un algoritmo que crea automáticamente un perfil de usuario jerárquico para representar el dominio y los intereses personales implícitos del usuario. Se representan el dominio y los intereses con una red conceptual de nodos enlazados entre sí. Esta red se realiza a través de relaciones respetando la topología de enlace (sinonimia, hiponimia e hiperonimia) definida en ontologías y el dominio de las jerarquías.

Modelo	Formas de representación	Ventajas	Desventajas	Aplicación
Basado en términos o palabras clave	Vectores numéricos ponderados	Más usado, de simple construcción	Polisemia	Recomendación de contenido, redes sociales, recuperación de información
Semántico	Representaciones semánticas, ontologías	Evita la polisemia, relaciona conceptos y semántica, representación explícita de conocimiento	Alta complejidad de cómputo y construcción	Motores de búsqueda, recuperación de información, turismo, comercio electrónico, <i>e-learning</i>
Jerárquico	Nodos de conceptos y relaciones	Tiene en cuenta sinonimia, hiponimia, hiperonimia	Alta complejidad	Recomendación de contenido

Tabla 2: Modelos de perfiles de preferencias de usuarios

La Tabla 2 resume los principales modelos de perfiles de usuarios. El estudio de este acápite evidencia que la mayoría de los modelos de perfiles son dependientes del dominio, por tanto, existe la necesidad de investigar más sobre los perfiles genéricos que puedan ser empleados en cualquier dominio. El lenguaje también juega un papel fundamental en el perfilado, sin embargo, todavía no son suficientes los sistemas que realicen este tipo de análisis. En consecuencia, se necesita continuar investigando para encontrar un mecanismo que pueda analizar el contenido en cualquier idioma y obtener un perfil más preciso.

Sugerencias para futuras investigaciones

Los resultados y las limitaciones de este estudio constituyen una base para investigaciones futuras sobre perfiles de preferencias de usuarios en Sistemas de Recuperación de Información. Para profundizar en estos temas se recomienda añadir otras bases de datos al estudio realizado (*Web of Science* y *Scopus*), de esta forma se abarcará más información y los resultados serán más exactos. Los estudios futuros pueden también centrar su

atención en el impacto real de contar con un generador de perfiles de usuario para la relevancia de los resultados de búsqueda en SRI cubanos como es el caso de Orión.

CONCLUSIONES

En esta investigación se analizó la literatura existente en la base de datos GS referente a la generación de perfiles de preferencias de usuarios en el periodo 2016-2022. El análisis se centró en la información almacenada, cómo extraerla y modelarla para construir un perfil de usuario. En el estudio realizado se determinó que actualmente es la región de Asia la mayor productora de publicaciones sobre el tema.

La revisión sistemática realizada corroboró la importancia de los perfiles de preferencias de usuarios en los servicios de personalización. Los resultados y las limitaciones de este estudio constituyen una base para investigaciones futuras sobre perfiles de preferencias de usuarios en Sistemas de Recuperación de Información.

REFERENCIAS BIBLIOGRÁFICAS

1. Abdeljaber, H. (2018). Profile-Based Semantic Method using Heuristics for Web Search Personalization. *International Journal of Advanced Computer Science and Applications*, 9. <https://doi.org/10.14569/IJACSA.2018.090926>
2. Bhavadharani, M., Ramkumar, M. P., & GSR, E. S. (2019). Information Retrieval in Search Engines Using Pseudo Relevance Feedback Mechanism. *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 1-5.
3. Calvopiña, K., & Pérez, V. (2018). *GUÍA DE COMPRAS EN ECUADOR APLICADA AL MODELO B2C EN BASE A LAS PREFERENCIAS EN LAS REDES SOCIALES* [Tesis de grado de Ingeniero de Sistemas, Universidad de las
4. Costanzo, L. L., Deldjoo, Y., Dacrema, M. F., Schedl, M., & Cremonesi, P. (2019). Towards Evaluating User Profiling Methods Based on Explicit Ratings on Item Features. *ArXiv Preprint ArXiv:1908.11055*.
5. Gao, G. (2019). A Survey of User Profiles Methods. *Data Analysis and Knowledge Discovery*, 3(3), Art. 3. <https://doi.org/10.11925/infotech.2096-3467.2018.0784>
6. García, L. S. (2021). *Sistema predictivo basado en el modelo web para la predicción de preferencias de usuario en el sitio web foros Perú* [Tesis Doctoral, UNIVERSIDAD SEÑOR DE SIPÁN]. <http://repositorio.uss.edu.pe/handle/20.500.12802/8225>
7. Rodríguez, N., Campos, P., & Cantador, I. (2017). *Recomendación contextualizada usando una ontología de contexto genérica y de gran escala construida semiautomáticamente a partir de DBpedia* [Tesis de Maestría, Universidad Autónoma de Madrid]. https://repositorio.uam.es/bitstream/handle/10486/681054/Rodriguez_Artigot_Nicolas_tfm.pdf?sequence=1
8. Rodríguez, P. A., Duque, N. D., & Ovalle, D. A. (2016). Método Híbrido de Recomendación Adaptativa de Objetos de Aprendizaje basado en perfiles de Usuario. *Formación universitaria*, 9(4), Art. 4.
9. Safi, H., Jaoua, M., & Belguith, L. H. (2017). Learning-to-Rank for Hybrid User Profiles. *Research in Computing Science*, 132, 107-127.
10. Viltres, H., Leyva, P. R., Febles, J. P., & Sentí, V. E. (2018). MODELO PARA LA RECUPERACIÓN DE INFORMACIÓN CON ANOTACIÓN SEMÁNTICA MODEL FOR THE RECOVERY OF INFORMATION WITH SEMANTIC ANNOTATION. *Informática 2018*. "VIII Congreso Internacional de Tecnologías y Contenidos Multimedia", Cuba.