

Tipo de artículo: Artículo de revisión
Temática: Inteligencia Artificial
Recibido: 17/11/2018 | Aceptado: 05/04/2018

Análisis de tendencias en la personalización de los resultados en buscadores web

Analysis of trends in the customization of results in web search engines

Eric Bárbaro Utrera Sust^{1*}, Alfredo Simón Cuevas², José A. Olivás³

¹ Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.:19370. ebutrera@uci.cu

²Universidad Tecnológica de La Habana “José Antonio Echeverría”, Cujae. Facultad de Informática. La Habana, Cuba. asimon@ceis.cujae.edu.cu

³Universidad de Castilla La Mancha, JoseAngel.Olivas@uclm.es.

***Autor para correspondencia:** ebutrera@uci.cu

Resumen

La web ha crecido exponencialmente en los últimos años. Una característica de los buscadores es que para una misma consulta hecha por dos usuarios diferentes devuelve los mismos resultados, por lo que la necesidad de personalizar los resultados de búsqueda se ha hecho cada vez más necesario. En este artículo se presenta una revisión de Modelos de Personalización para Buscadores Web (MPBW), con un enfoque en la búsqueda personalizada. La investigación proporciona una revisión de las etapas involucradas en la construcción y evaluación de los MPBW como son: modelación de las características de los usuarios y las relaciones entre estos, recopilación de la información sobre los usuarios del sistema, representación de la información, trabajo multi-idioma, confiabilidad de los datos, creación de perfiles y uso de agentes. Con base al análisis realizado, el documento concluye poniendo en evidencia los desafíos y las futuras direcciones de investigación en el campo del MPBW.

Palabras clave: buscadores web, búsqueda personalizada, modelos de personalización para buscadores web, recuperación de información

Abstract

Web search engines help users find specific information within the large amount of information resources available on the Web. A feature of web search engines is that for the same query made by two different users, they show the same

results, so the need to customize search results has become increasingly necessary. This article presents a review of Personalization Models for Web Search Engines (PMWS), with a focus on personalized search. The research provides a vision of the stages involved in the construction and evaluation of the MPWB, such as: modeling of users' characteristics and the relationships among them, compilation of the information about the users of the system, representation of the information, multi-language performance, reliability of data, creation of profiles and use of agents. Based on the analysis carried out, this paper concludes by highlighting the challenges and future directions of research in the field of PMWS.

Keywords: *information retrieval, personalized search, personalization models for web search engines, web search engines.*

Introducción

El constante crecimiento de la información disponible en internet ha incentivado el desarrollo de herramientas que permitan centralizar la mayoría de los datos de la web en un solo lugar. Estas herramientas nombradas buscadores, son capaces de, a partir de un criterio de búsqueda introducido por el usuario, mostrar resultados que pueden satisfacerle o no su necesidad de información. Una de las interrogantes que se plantean permanentemente los desarrollos de buscadores es ¿cómo disminuir la información que no es de interés para el usuario en los resultados de las búsquedas? En respuesta esta interrogante, varios investigadores han reportado el uso de técnicas de Minería de Datos, Minería de Textos y Minería Web, pero la solución no solo se queda aquí, se expande a nuevos campos como el de la Personalización de la Web (PW) (SINGH, 2017).

Este nuevo campo se basa en conocer las características del usuario, creando un modelo que lo caracterice, lo que posibilita personalizar los resultados de búsqueda. Un buscador devuelve el mismo resultado a un usuario biólogo y otro informático que introducen la misma consulta “ratón”, debido a que un buscador para devolver resultados se basa en las palabras claves de la consulta y los documentos. Sin embargo, si un buscador es capaz de caracterizar al usuario que introduce la consulta en este caso basándose en su profesión, es capaz también de personalizarle los resultados. Varios estudios (DOU, 2009); (SHAPIRA, 2011); (MATTHIJS, 2011); (D. ZHOU, et al., 2016) han demostrado que personalizando los resultados de búsqueda se mejora la relevancia de los buscadores. En este artículo se hace un estudio de varios Modelos de Personalización para Buscadores Web (MPBW) propuestos por investigadores, teniendo en cuenta como modelan de las características de los usuarios y las relaciones entre ellos, la forma de recopilar las

preferencias de los usuarios, como representan la información, como tienen en cuenta otros idiomas para la personalización, como manejan la privacidad de los datos, la creación de perfiles y el trabajo con agentes. En otros trabajos como (GHORAB, 2013) se hacen revisiones a modelos de personalización. Teniendo en cuenta que se hizo en 2013 y que no abordan características que son de interés para esta investigación se hace este estudio.

El resto de este artículo está estructurado de la siguiente manera: En la Sección II se explicará la metodología utilizada para seleccionar los trabajos significativos en cuanto a MPBW. La Sección III describe los fundamentos sobre los modelos de personalización enfocándose en: extracción y análisis de las preferencias, representación de la información, trabajo multi-idioma, confiabilidad de los datos, creación de perfiles y uso de agentes. En la Sección IV se hace un resumen general del estudio según las características de cada modelo y la Sección V concluye la investigación realizada y propone nuevas áreas de investigación.

Materiales y métodos

Se realizó un estudio inicial para determinar los temas y términos más representativos en el campo de los MPBW. Los métodos de investigación empleados fueron el analítico-sintético e histórico-lógico. El método analítico-sintético se empleó para analizar la bibliografía referente a modelos de este tipo, sintetizando posteriormente particularidades que los caracterizaban. El método histórico-lógico se utilizó para determinar las distintas etapas de los modelos descritos y la evolución de estos en cuanto a su novedad. Se seleccionaron trabajos sobre MPBW de revistas y conferencias mostradas en la Tabla 1.

Tabla 1. Revistas con publicaciones en el campo de los MPBW.

Fuente	Cantidad
IEEE	13
ACM	7
International Journal Of Advanced Computer Science And Applications	4
Springer	3
International Journal Of Advanced Research In Computer Science	2
Advances In Computational Sciences And Technology	1
International Conference On Recents Advanced In Engineering Science And Management	1
En International Conference On Information And Communication Technology For Intelligent Systems. Springer, Cham	1
International Journal Of Pure And Applied Mathematics	1
Software: Practice And Experience	1
Proceedings Of The 2016 Conference On Empirical Methods In Natural Language Processing	1
ARNP Journal Of Engineering And Applied Sciences	1
International Journal Of Research In Computer Applications And Robotics	1

I.J. Mathematical Sciences And Computing	1
International Research Journal Of Engineering And Technology (IRJET)	1
World Academy Of Science, Engineering And Technology, International Journal Of Computer, Electrical, Automation, Control And Information Engineering	1
International Journal Of Applied Engineering Research	1
International Journal Of Computer & Mathematical Sciences	1
International Journal Of Soft Computing	1
The Journal Of Web Science	1
Communications In Computer And Information Science	1
International Journal Of Service Science, Management, Engineering, And Technology (IJSSMET)	1
Information Sciences	1
International Journal Of Web & Semantic Technology (IJWEST)	1
International Journal Of Innovative Research In Advanced Engineering (IJIRAE)	1
International Journal Of Innovative Research In Computer And Communication Engineering	1
Journal Of Network And Computer Applications	1
International Journal Of Advance Research In Computer Science And Management Studies	1
International Journal Of Industrial Organization	1

Se hizo un análisis también de estos artículos por años según su evolución en el tiempo. En la siguiente gráfica se muestra la cantidad de artículos por año a partir de los trabajos seleccionados.

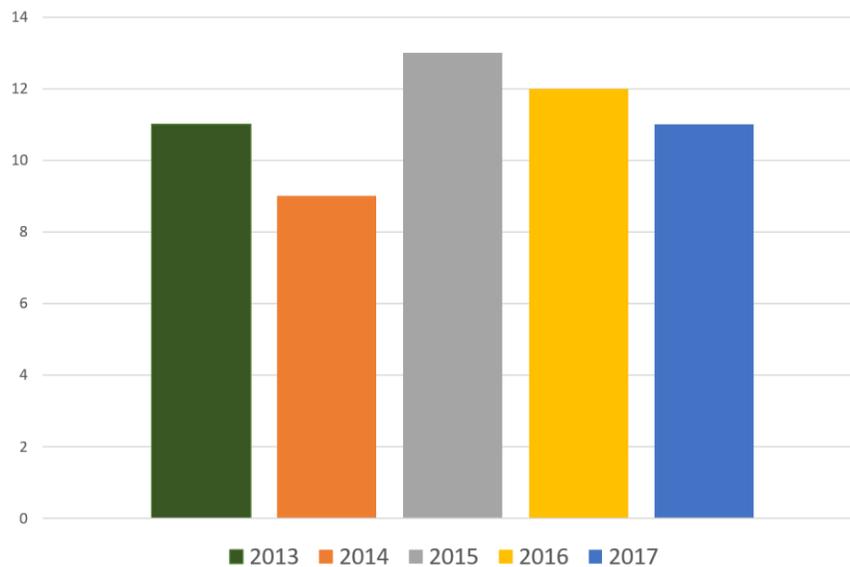


Figura 1. Comportamiento de investigaciones sobre MPBW en el tiempo.

El 99% de los artículos son en idioma inglés teniendo en cuenta que es el idioma en el que más se publica en materia científica (BASTURKMEN, 2016). Se revisaron trabajos a partir de los más citados según Google Scholar en este

campo sobresaliendo trabajos de autores como Susan Dumais citada por 57958, Bracha Shapira citada por 5359, Dong Zhou citado por 388 y Kamlesh M. Makwana citado por 68 autores,

Modelos de Personalización en Buscadores Web (MPBW)

Los modelos de personalización en buscadores web tienen como objetivo brindar resultados personalizados a los usuarios, disminuyendo los documentos irrelevantes mostrados al usuario (SEKHARBABU, 2016; SINGH, 2017). La idea central de estos modelos es caracterizar a los usuarios que interactúan con el sistema a partir de la generación de un perfil que esté en correspondencia con sus gustos e intereses (D. ZHOU, et al., 2016); (LOVARAJU, 2017); (MAKWANA, 2017).

Estos modelos han ganado auge en los últimos tiempos y han sido estudiados y trabajados por múltiples autores para dar solución a problemas determinados (BABEKR, 2013; BIBI, 2014; BOSTAN, 2015; DUMAIS, 2016; GAO, 2013; GHORAB, 2013; HANNAK, 2013; JOHNSON, 2016; LI, 2014; LOVARAJU, 2017; MAKVANA, 2014; MAKWANA, 2017; MOR, 2015; PREETHA, 2014; RAJMANE, 2016; ROPHIE, 2016; SEKHARBABU, 2016; SHAFIQ, 2015; SHARDA, 2017; SHARMA, 2017; SINGH, 2017; USTINOVSKIY, 2013; VANITHA, 2013; WANG, 2017; ZHANG, 2017; D. ZHOU, et al., 2016; L. Zhou, 2015).

1 Personalización en Buscadores

A continuación, se hace un análisis de varios trabajos orientados a la personalización de los resultados en buscadores web.

1.1 Extracción y análisis de las preferencias

En (USTINOVSKIY, 2013) se aborda el problema de la personalización de la búsqueda web para nuevos usuarios. En esta investigación se propone realizar personalizaciones a **corto plazo**¹ ya que se parte de la idea de que el historial de búsqueda a **largo plazo**² es escaso o está completamente ausente para los usuarios. En esta propuesta se extraen varias estadísticas de las sesiones de navegación de otros usuarios para mejorar la clasificación del usuario actual. Tienen en cuenta que las páginas web visitadas por un usuario antes de formular una consulta Q en la misma sesión se denominan contexto de exploración y las consultas emitidas antes de Q, respectivamente, se denominan contexto de búsqueda.

¹ Los intereses a corto plazo son aquellos que se van readaptando con el tiempo (DING, 2017; L. Zhou, 2015).

² Los intereses a largo plazo son constantes en el tiempo y el usuario es el encargado de definirlos en el momento del registro (DING, 2017; L. Zhou, 2015).

Aquí tienen en cuenta las características de la consulta Q , la utilidad de la sesión de navegación B como contexto de consulta Q y la proximidad basada en clics entre un usuario u_1 y usuario u_2 para crear los perfiles.

En (VANITHA, 2013) se propone una idea novedosa de combinar dos métodos cruciales: el perfil del usuario y los clics del usuario para mejorar la búsqueda basada en la web y aumentar el nivel de personalización. Cada uno de estos métodos está diseñado como un módulo separado y el resultado de estos dos módulos se utiliza para encontrar las páginas web de manera más personalizadas. En este trabajo se establece una estrategia que se basa en asignar la consulta del usuario a un conjunto de categorías que representan la intención de búsqueda del usuario. Una categoría en un perfil de usuario sería un vector de términos ponderados, en el cual un alto peso de un término indica que el término tiene una gran importancia en esa categoría para el usuario. En (BIBI, 2014) se propone un perfil con un enfoque híbrido basado en conceptos y documentos. Se propone un marco de trabajo que extrae datos de los enlaces a los que los usuarios le dan clic en el navegador. Se analizan, además, las consultas y los documentos devueltos para cada una de ellas. En (MAKVANA, 2014) se personaliza el resultado de la búsqueda web mediante la reformulación de consultas y el perfil del usuario. Para reformular la consulta se basan en identificar su ambigüedad recuperando términos útiles mediante el análisis de los términos de búsqueda anteriores del usuario y finalmente anexan la palabra clave más relevante a la consulta del usuario. En (THANGARAJ, 2014) se usan dos métodos generales para descubrir el interés del usuario, la retroalimentación aparente y la retroalimentación connotativa. En la retroalimentación aparente, el usuario puede ingresar los datos de interés personal o evaluación al trabajo actual. En la retroalimentación connotativa el sistema puede obtener información de interés del usuario a través del seguimiento del comportamiento y la operación del usuario, para ello se basan en el registro histórico, y el comportamiento del usuario. El algoritmo propuesto asigna un puntaje que mide la calidad y la relevancia de un conjunto seleccionado de páginas en función de su URL para una consulta de usuario determinada. Luego se crea una tabla de vectores bidimensionales específica de la consulta, denominada tabla de vectores relacionados, y realiza el análisis de URL. En (PREETHA, 2014) se crean sesiones de retroalimentación a partir de un conjunto de consultas web para inferir resultados. Tanto las URL clicadas como las cliqueadas antes del último clic se consideran como retroalimentaciones implícitas del usuario y se tienen en cuenta para construir sesiones de retroalimentación. En (KUMAR, 2014) se propone un marco para construir un perfil de usuario mejorado mediante el uso del historial de navegación del usuario y enriqueciéndolo usando una base de conocimiento. La forma de modelar el usuario en este trabajo es a través de un vector documento-preferencia, donde el peso de cada documento representa el interés del usuario por esa categoría. En (BOSTAN, 2015) se basan el historial de búsqueda de cada usuario en un marco de tiempo específico y la utilizan para ofrecer resultados personalizables y más eficientes para el usuario. En (SINGH, 2015) se propone un marco basado en la web semántica y la tecnología de

agente para la recuperación de información personalizada de la web en el ámbito agrícola. En este trabajo presentan 4 agentes encargados de identificar el contexto y modelar el perfil del usuario a través de una ontología. En (SHAFIQ, 2015) se plantea que las consultas introducidas en un buscador por los usuarios puede tener una fuerte correlación con la información relevante en sus redes sociales. Para encontrar los intereses personales y los contextos sociales extraen la información de las actividades de los usuarios en sus redes sociales y usan esa información para volver a clasificar los resultados de un motor de búsqueda. En (MOR, 2015) se investigan técnicas prácticas para ofuscar el perfil de un usuario de una manera que preserve la privacidad del usuario y, a la vez, permita que el servidor de personalización personifique los resultados de una manera útil. Para esto utilizan dos técnicas bien conocidas para la ofuscación de perfil: la generalización que comparte elementos en el perfil de un usuario solo en una granularidad aproximada (por ejemplo, categoría de sitios web visitados con frecuencia, en lugar de sitios reales) y adición de ruido que agrega elementos falsos al perfil para ocultar los elementos reales. Hacen uso de los registros de búsqueda de un motor de búsqueda popular para cuantificar las compensaciones. En (VERMA, 2016) se propone un motor de búsqueda multi-agente semántico que procesa la consulta recuperada de los motores de búsqueda tradicionales y analiza estos resultados de acuerdo con la prioridad del usuario con la ayuda de múltiples agentes. Este motor de búsqueda es beneficioso porque tiene un módulo de dominio de ontología que se utiliza para representar la relación entre las preferencias del usuario y los resultados de búsqueda producidos. El nivel de personalización disminuye del Nivel 3 al 0. Dependiendo del requerimiento del usuario, la personalización se puede lograr utilizando ambos (personalización del lado del cliente y personalización del lado del servidor), ya sea o ninguno. En (D. ZHOU, et al., 2016) se aborda el entorno en el que un usuario tiene solo una cantidad limitada de información de uso. Construyen perfiles de usuario mejorados a partir de un conjunto de anotaciones y recursos que los usuarios han marcado, junto con una base de conocimiento externa construida de acuerdo con los historiales de uso. Presentan dos modelos probabilísticos de temas latentes para incorporar simultáneamente anotaciones sociales, documentos y la base de conocimiento externa. La estrategia de búsqueda web la logran mediante la expansión de consultas sociales personalizadas. Además, presentan un modelo de expansión de consulta tópica para mejorar la búsqueda mediante la utilización de perfiles de usuario individuales. En (SEKHARBABU, 2016) se presenta una propuesta para mejorar la calidad de las búsquedas manteniendo el control de privacidad del usuario. La calidad en la búsqueda se obtiene utilizando el perfil del usuario creado mediante el uso del historial y las búsquedas del usuario. La pérdida de datos confidenciales debe controlarse durante el proceso de procesamiento de consultas. En el trabajo de (JOHNSON, 2016) se implementan dos métodos, uno basado en el dominio el cual hace inferencias sobre los intereses del usuario y otro método de similitud basado en la taxonomía, utilizado para refinar el algoritmo de coincidencia de elementos del usuario, mejorando los resultados generales. El recomendador

propuesto es independiente del dominio, se implementa como un servicio web y utiliza métodos de recopilación de comentarios tanto explícitos como implícitos para obtener información sobre los intereses del usuario. En (WANG, 2017) se propone un método personalizado para fusionar resultados de los metabuscadores de acuerdo con una variedad de factores, incluido el interés del usuario, distribución, el número total de motores de búsqueda, el número de resultados que devuelve cada motor, la posición de clasificación del documento en cada motor de búsqueda y el número de motores de búsqueda que devolvieron el documento. En (MAKWANA, 2017) se crea un enfoque para mitigar el arranque en frío, problema que es común en los sistemas de recomendación. En este trabajo crean un enfoque que se divide en tres fases. En la primera fase, se analiza el archivo de registro web que se ha generado en formato predefinido a través de los clics del usuario. Una base de conocimiento se incorporará en forma de ontología mediante el análisis del archivo de registro web. Al mismo tiempo, el usuario dará alguna información explícita al sistema. En la segunda fase, se crean grupos de usuarios similares utilizando la técnica de agrupamiento de c-means difusa. Tanto los datos de la primera como de la segunda fase se utilizan como entrada del sistema. En la última fase, la asignación de ontologías se realiza en función de la técnica de filtrado colaborativo de elemento a elemento.

A continuación, en la Tabla 2 se presenta un resumen de esta sección:

Tabla 2. Resumen sobre extracción y análisis de las preferencias en MPBW.

Trabajos	Enfoque de recopilación de información	Fuentes de preferencias	Técnicas utilizada
(USTINOVSKIY, 2013)	Implícito	Consultas, documentos consultados o fragmentos de documentos consultados	• Método de re-ranking
(VANITHA, 2013)	Implícito	Consultas, documentos consultados o fragmentos de documentos consultados	• Clasificador bayesiano
(BIBI, 2014)	Implícito	Conceptos de los documentos, consultas, documentos consultados	• Similitud entre conceptos en una ontología
(MAKVANA, 2014)	Explícito	Consultas, términos de búsqueda anteriores	• Frecuencia de un término. • Técnica basada en palabras claves.
(THANGARAJ, 2014)	Explícito e implícito	Datos de interés personal o evaluación al resultado. Información de interés del usuario a través del seguimiento del comportamiento y las operaciones del usuario.	• Tecnología de Agentes

(PREETHA, 2014)	Explícito e implícito	Consultas, documentos consultados	• Método basado en inferencias.
(KUMAR, 2014)	Implícito	Historial de navegación del usuario	• AlchemyAPI
(BOSTAN, 2015)	Implícito	Historial de búsqueda	• Agrupamiento K-Means • Modelo TF.IDF
(SINGH, 2015)	Implícito	Consultas, documentos	• Tecnología de Agentes
(SHAFIQ, 2015)	Explícito	Consultas, redes sociales	• Estrategias de análisis de redes sociales. • Técnica de cascada fraccional
(MOR, 2015)	Implícito	Documentos consultados	• Grano fino (Basado en URL) • Grano grueso (Basado en intereses)
(VERMA, 2016)	Explícito e implícito	Consultas, documentos	• Tecnología de Agentes
(D. ZHOU, et al., 2016)	Implícito	Historiales de navegación, anotaciones externas, documentos	• Mezcla mejorada • Mejora separada
(SEKHARBABU, 2016)	Explícito e implícito	Consultas, historial de navegación	• Algoritmos Greedy
(JOHNSON, 2016)	Explícito e implícito	Consultas, documentos consultados	• Método basado en inferencias. • Similitud en taxonomías • Algoritmos de stemming
(WANG, 2017)	Implícito	Consultas, documentos presentes en buscadores	• Segmentación de las palabras. • Naive Bayesian
(MAKWANA, 2017)	Explícito	Consultas, documentos consultados	• Agrupamiento difuso C-means

1.2 Representación de la información

En (USTINOVSKIY, 2013) se utiliza una base de conocimientos sobre la experiencia de navegación reciente de otros usuarios para mejorar la experiencia de búsqueda de un nuevo usuario. En (BIBI, 2014) todos los conceptos obtenidos desde un clic son organizados en forma de árbol usando las diferentes relaciones que existen entre los conceptos (ontología). A medida que se da un clic a un determinado documento, incrementan los pesos asociados a los conceptos del documento que se encuentran en la ontología. En (MAKVANA, 2014) se representa una base de conocimiento en RDF³ con conceptos utilizados para desambiguar la consulta introducida por el usuario. En (THANGARAJ, 2014) se hace uso de una base de conocimiento donde se incluye la información personalizada del usuario transmitida por uno de sus agentes, como puede ser el comportamiento de interés del usuario y el historial de búsqueda. Este trabajo también hace uso de ontologías de dominios para hacer expansión de la consulta y enriquecerla semánticamente. En (MALA, 2014) las preferencias de cada usuario se ordenan en un modelo basado en ontología y cada perfil de usuario se clasifica con el uso de multifacético para futuros resultados de búsqueda. El resultado de búsqueda se puede clasificar en

³ RDF (Resource Description Framework), Marco de Descripción de Recursos (en español).

conceptos de ubicación y contenido en función de la importancia de su información. En (KUMAR, 2014) se utiliza una base de conocimientos del dominio que permite almacenar los diferentes dominios y categorías que se están analizando. En (SINGH, 2015) se hace uso de una ontología que representa el dominio de la agricultura. La ontología es utilizada por los 4 agentes propuestos para personalizar la experiencia de los usuarios en la agricultura, identificando el interés del usuario a través del comportamiento de uso de la web. En el trabajo de (VERMA, 2016) el motor de búsqueda que proponen se basa en una ontología de dominio que se crea utilizando fases de desarrollo de ontología para garantizar la representación jerárquica de los resultados. En (ROGUSHINA, 2016) se propone un modelo ontológico de búsqueda semántica que incluye una ontología de dominio. Esta ontología describe el área de las necesidades de información del usuario basándose en una ontología léxica de dominio. La misma contiene información sobre fragmentos de lenguaje natural utilizados para establecer enlaces entre elementos de documentos de lenguaje natural y términos de ontología. Por otra parte, proponen un tesoro de tareas, un conjunto de consultas relacionadas con necesidades de información semánticamente similares y un conjunto de pares (referencia al documento buscado, clasificación del resultado). Además, incluyen una clase usuario que divide la información de los usuarios en varios grupos y un recomendador que muestra resultados al usuario propuestos por el Sistema de Recuperación de Información de manera proactiva. En el trabajo de (JOHNSON, 2016) se representa la información como una ontología para cada sitio web, donde los conceptos y términos importantes son extraídos de los documentos. De acuerdo con la similitud semántica de los documentos web para agruparlos en diferentes temas semánticos, los diferentes temas implican diferentes preferencias. En el trabajo de (LOVARAJU, 2017) se propone un modelo de ontología para representar el conocimiento previo del usuario para la recopilación de información web personalizada. El modelo construye ontologías personalizadas por el usuario extrayendo conocimiento del mundo y descubriendo el conocimiento del usuario de los repositorios de instancias locales de los usuarios. Un método de minería ontológica multidimensional, de manera exhaustiva y específica, también se introduce para el descubrimiento de conocimiento de fondo del usuario. En (MAKWANA, 2017) se genera una base de conocimiento analizando los datos del registro de consultas. El registro de consulta se genera a partir de la búsqueda del usuario y la información de clic. La base de conocimiento presenta la taxonomía de consultas buscadas por todos los usuarios. El elemento raíz de cada etiqueta representa la clave de los usuarios individuales que lo distinguen de otros usuarios y también proporciona la privacidad a nivel del cliente.

A continuación, en la Tabla 3 se presenta un resumen de esta sección:

Tabla 3. Resumen sobre representación de información en MPBW

Autor	Representación de las preferencias	Tecnología utilizada para contribuir a la representación de las preferencias
--------------	---	---

(USTINOVSKIY, 2013)	Modelo basado en vectores	<ul style="list-style-type: none"> • Base de conocimiento (Almacenando la experiencia de navegación de los usuarios (acciones y documentos cliqueados))
(BIBI, 2014)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontología (Representando una estructura de árbol que simboliza la diferentes relaciones que existen entre los conceptos cliqueados)
(MAKVANA, 2014)	Modelo basado en vectores	<ul style="list-style-type: none"> • RDF (Representando un grafo con conceptos utilizados para desambiguar la consulta introducida por el usuario)
(THANGARAJ, 2014)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontologías (Representando relaciones que existen entre los conceptos) • Base de conocimiento (Almacenando la información personalizada)
(MALA, 2014)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontologías (Representando relaciones entre conceptos)
(KUMAR, 2014)	Modelo basado en vectores	<ul style="list-style-type: none"> • Base de conocimiento (Almacenando los diferentes dominios y categorías que se están analizando)
(SINGH, 2015)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontologías (Representando relaciones entre conceptos en el dominio de la agricultura)
(VERMA, 2016)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontologías (Representando relaciones entre conceptos en los dominios)
(ROGUSHINA, 2016)	Modelo semántico	<ul style="list-style-type: none"> • Ontologías (Representando relaciones entre conceptos)
(JOHNSON, 2016)	Modelo basado en vectores	<ul style="list-style-type: none"> • Ontologías (Representando relaciones entre los términos y conceptos de los documentos)
(LOVARAJU, 2017)	Modelo semántico	<ul style="list-style-type: none"> • Ontologías (Representando el conocimiento previo del usuario)
(MAKWANA, 2017)	Modelo basado en vectores	<ul style="list-style-type: none"> • Base de conocimiento (Almacenando los datos del registro de consultas)

1.3 Trabajo multi-idioma

El cambio de la web originalmente dominada por el idioma inglés a una red mundial realmente global ha generado una necesidad apremiante de desarrollar soluciones novedosas que aborden la diversidad de usuarios multilingües. En (GHORAB, 2014) evalúan la eficacia de la recuperación de algoritmos de personalización de búsqueda multilingüe con respecto a consultas de usuarios en inglés vs. no inglés. El estudio involucra a usuarios que provienen de diferentes entornos lingüísticos, interactuando con un sistema de búsqueda web que facilita el acceso a los resultados de búsqueda de múltiples idiomas. Los resultados de la evaluación muestran que los usuarios que no hablan inglés se benefician más del proceso de personalización de búsqueda que los usuarios de inglés. En (PHAM, 2016) se propone un enfoque léxico

basado en concordancia para tratar el multilingüismo en el proceso de recomendación y se muestra un experimento eficiente para un sistema de recomendación multilingüe en el dominio de la película. En (AMEUR, 2017) se presenta un sistema de recuperación de información basado en la personalización de resultados. Esta propuesta realiza reconocimiento del contexto teniendo en cuenta los contextos de usuario multilingües para modelar mejor los intereses de búsqueda de los usuarios. Los resultados experimentales demuestran la consistencia y adecuación de la propuesta.

1.4 Privacidad de los datos

En (SHOU, 2014) se estudia la protección de la privacidad en las aplicaciones para la búsqueda web personalizada y utilizan dos algoritmos Greedy DP y GreedyIL, para la generalización en tiempo de ejecución. Proponen un marco de búsqueda web personalizado que preserva la privacidad del usuario y que puede generalizar los perfiles de cada consulta de acuerdo con los requisitos de privacidad especificados por el usuario. En (DESAI, 2015) se propone también un marco de búsqueda web personalizado que puede generalizar de forma adaptativa los perfiles mediante consultas, respetando los requisitos de privacidad especificados por el usuario. Esta generalización tiene como objetivo lograr un equilibrio entre dos indicadores predictivos que evalúan la utilidad de la personalización y el riesgo de privacidad de exponer el perfil generalizado. En este trabajo utilizan también los algoritmos GreedyDP y GreedyIL, para soportar el perfil en tiempo de ejecución. En (YANG, 2015) se destaca la necesidad de explorar los parámetros de consulta de búsqueda y determinar su impacto en la personalización. Este es un primer paso para explorar los mecanismos de recopilación de datos personales y cómo la búsqueda personalizada utiliza datos personales, lo que posteriormente afecta la privacidad de la información de los usuarios. En (MOR, 2015) se propone las cookies de Bloom que codifican el perfil de un usuario de forma compacta y que preserva la privacidad, impidiendo que los servicios en línea lo utilicen con fines de personalización. En este trabajo plantean que la generalización del perfil significativamente perjudica la personalización y no protege a los usuarios de un servidor que vincula sesiones de usuario a lo largo del tiempo. En (RAKESH, 2016) se analizan formas en que se puede mejorar la privacidad para que los usuarios puedan sentirse más cómodos con la publicación de sus datos personales a fin de recibir resultados de búsqueda más precisos. En este trabajo se ayuda al usuario a proporcionar la información sobre si personalizar el perfil para la consulta dada. Si el usuario cree que la consulta es personal, la consulta se enviará al servidor con información limitada sobre el perfil del usuario. Esto evitará el riesgo que pueda ocurrir en la privacidad del usuario. El marco propuesto puede generalizar los perfiles para cada consulta de acuerdo con los requisitos de privacidad especificados por el usuario. Aquí se usarán dos algoritmos llamados GreedyDP y GreedyIL para la generalización de tiempo de ejecución. Mientras que el primer algoritmo intenta maximizar el poder de discriminación, el otro intenta minimizar la pérdida de información. En (SEKHARBABU, 2016)

se desarrolla un componente para proteger la privacidad el cual genera un perfil en línea que se aplica en un proxy de búsqueda que se ejecuta en una máquina cliente. Este proxy tendrá el perfil de usuario jerárquico y los requisitos de privacidad personalizados.

A continuación, en la Tabla 4 se presenta un resumen de esta sección:

Tabla 4. Resumen sobre privacidad de los datos en MPBW.

Autores	Tecnologías utilizadas
(SHOU, 2014)	Algoritmo Greedy DP, Algoritmo GreedyIL
(DESAI, 2015)	Algoritmo Greedy DP, Algoritmo GreedyIL
(MOR, 2015)	Cookies de Bloom
(RAKESH, 2016)	Algoritmo Greedy DP, Algoritmo GreedyIL
(SEKHARBABU, 2016)	Algoritmo Greedy DP, Algoritmo GreedyIL

1.5 Uso de agentes

En (THANGARAJ, 2014) se hace uso de la colaboración de múltiples agentes para llevar a cabo la recuperación de información de acuerdo con el conocimiento de interés del usuario. En el proceso de recuperación de información personal, la precisión y la calidad dependen del grado de veracidad que el usuario maestro del sistema le interese. Presentan un agente de usuario encargado de construir el modelo de interés del usuario de acuerdo con el registro del historial de navegación del usuario y los datos de registro. Un agente de extracción semántica encargado de encontrar las características semánticas en las consultas de los usuarios. Utiliza datos de otros agentes y tecnologías de ontología para analizar la relación de asociación en las consultas y el documento de los usuarios y así extraer características semánticas. Un agente de búsqueda semántica el cual es el responsable de buscar y recuperar resultados relevantes. Un agente de filtrado capaz de realizar una rápida coincidencia y búsqueda de contenido y un último agente de clasificación personalizado el cual es el centro de las tomas de decisiones del sistema de recuperación de información personalizado. En (SINGH, 2015) se propone un marco basado en agentes semánticos e inteligentes para la recuperación de información personalizada en la agricultura. En este trabajo presentan 4 agentes, un agente agricultor encargado de personalizar la interfaz web de los agricultores, un agente investigador responsable de identificar los intereses y preferencias de los investigadores, un agente académico que se encarga de identificar los intereses y personalizaciones de los usuarios que no son investigadores y un último agente comerciante responsable de personalizar la experiencia comercial de los agricultores. Estos agentes hacen uso de una ontología de dominio para su funcionamiento. En el trabajo de (VERMA, 2016) se propone una arquitectura multi-agente para producir resultados personalizados. El

propósito de la investigación es proporcionar una plataforma para búsqueda personalizada específica de dominio. La arquitectura que proponen usa la personalización del lado del cliente y del lado del servidor para proporcionar al usuario resultados más precisos. La arquitectura del motor de búsqueda multi-agente usa el concepto de descriptores semánticos para adquirir conocimiento sobre el dominio dado y conducir a resultados de búsqueda personalizados. Las descripciones semánticas se representan como un gráfico de red que mantiene la relación entre un problema determinado en forma de jerarquía. En este trabajo presentan un agente de interfaz (agente de usuario) capaz de interactuar con el usuario para proporcionar requisitos y mostrar los resultados. Tiene un módulo de interfaz que contiene un método para la comunicación entre agentes. Un agente facilitador (agente de administración) que activa diferentes agentes, recibe preguntas del agente de interfaz y puede tomar la ayuda de un grupo de agentes para resolver esas preguntas. Un agente de recursos (Agente de datos) que mantiene información de metadatos sobre las fuentes de datos. Este agente además genera consultas basadas en la solicitud del usuario y envía sus resultados al agente de usuario. Un agente de minería que implementa técnicas y algoritmos de minería de datos. Un agente de resultados que observa agentes mineros y otros resultados de ellos. Después de obtener los resultados, este agente muestra los resultados al agente de usuario al integrarse con el agente de administrador. Un agente de corretaje que puede enviar una respuesta a la consulta de un agente con el nombre y la ontología del agente respectivo. El último agente es un agente de ontología que mantiene y proporciona conocimiento de ontología para resolver consultas relacionadas con ontología. A continuación, en la Tabla 5 se presenta un resumen de esta sección:

Tabla 5. Resumen trabajo con agentes en MPBW

Autores	Fuente de conocimiento	Algoritmo de aprendizaje	Tipos de agentes
(THANGARAJ, 2014)	Ontologías	Re-ranking	<ul style="list-style-type: none"> • Agente de usuario • Agente de extracción de semántica. • Agente de búsqueda semántica. • Agente de filtrado • Agente de ranking personalizado
(SINGH, 2015)	Ontologías	-	<ul style="list-style-type: none"> • Agente agricultor • Agente investigador • Agente académico • Agente comercial
(VERMA, 2016)	Ontologías	Técnicas de minería de datos	<ul style="list-style-type: none"> • Agente de usuario • Agente facilitador • Agente de recurso • Agente de minería • Agente de resultados • Agente corredor

Conclusiones

En este artículo se ha analizado una serie de investigaciones en el campo de la personalización de buscadores web teniendo en cuenta la modelación de las características de los usuarios y las relaciones entre ellos, la forma de recopilar las preferencias de los usuarios, representación de la información, manejo de idiomas, medidas utilizadas para la privacidad de los datos, creación de perfiles y el trabajo con agentes. Como resultado del análisis de este trabajo se puede decir que en este campo el enfoque de recopilación de información que más predomina es el implícito y la información más analizada para generar los perfiles de los usuarios son las consultas y los documentos consultados. Se puede, además, observar cómo se ha convertido en una tendencia, a partir del desarrollo de tecnologías de la web semántica, el uso de soluciones utilizando ontologías y grafos RDF. La gran mayoría de los trabajos representan la relación entre los usuarios basándose en el modelo de vectores y la tecnología más utilizada para contribuir al modelado son las ontologías de dominio. En ningún trabajo estudiado se construyen ontologías generalizadas integral, siendo esto esencial para los buscadores web debido a la cantidad de dominios que manejan. Por otra parte, teniendo en cuenta que la comunidad web ha llegado a un nivel en el que la multilingüidad global se está convirtiendo en un aspecto cada vez más importante de la interacción diaria de los usuarios con la información en la Web, aún en el campo de la personalización se debe considerar cómo puede adaptarse, ya que hay pocos trabajos que lo tienen en cuenta. Insuficientes soluciones plantean medidas para la privacidad de los datos y los algoritmos más utilizados son GreedyDP y GreedyIL. En los modelos de personalización para buscadores web aún el trabajo con agentes es escaso, utilizados ampliamente por su capacidad de ajustar su rendimiento a una preferencia individual del usuario, aprendiendo de sus acciones y modelando perfiles apropiados. Finalmente, el modelado preciso de los usuarios sin duda ayudará a proporcionar mejores y más efectivos sistemas de personalización en la Web.

Referencias

- AMEUR, M. SEGHIR H., et al. *Toward a Context-Aware Multilingual Personalized Search*. International Conference on Arabic Language Processing. Springer, Cham, 2017. p. 175-190.
- BABEK, S. T. F., KHALED M., et al. *Personalized semantic retrieval and summarization of web based documents*. International Journal of Advanced Computer Science and Applications, 2013, vol. 4, no 1, p. 177-186.
- BASTURKMEN, H. *English as a Scientific and Research Language*. Debates and Discourses: English in Europe, 2016, vol. 2, no 32, p. 245-248.
- BIBI, T., DIXIT, P., et al. *Web search personalization using machine learning techniques*. En Advance Computing Conference (IACC), IEEE International. IEEE, 2014, p. 1296-1299.

BOSTAN, S., GHASEMZADEH, G. *Personalization of Search Engines, Based-on Comparative Analysis of User Behavior*. Journal of Advances in Computer Research, 2015, vol. 6, no 2, p. 65-72.

DESAI, A., CHANDRE, P. *Privacy Preservation in Personalized Web Search*. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2015, vol 2, no 10.

DING, Y., et al. *Exploiting long- term and short- term preferences and RFID trajectories in shop recommendation*. Software: Practice and Experience, 2017, vol. 47, no 6, p. 849-865.

DOU, Z., et al. *Evaluating the effectiveness of personalized web search*. IEEE Transactions on Knowledge and Data Engineering, 2009, vol. 21, no 8, p. 1178-1190.

DUMAIS, S. T. *Personalized Search: Potential and Pitfalls*. CIKM., 2016, p. 689.

GAO, Q., YOUNG, I. *A multi-agent personalized ontology profile based query refinement approach for information retrieval*. Control, Automation and Systems (ICCAS), 2013 13th International Conference on. IEEE, 2013, p. 543-547.

GHORAB, M. R., et al. *Personalised Information Retrieval: survey and classification*. Springer, 2013, vol. 23, no 4, p. 381-443.

GHORAB, M. R., et al. *Does Personalization Benefit Everyone in the Same Way? Multilingual Search Personalization for English vs. Non-English Users*. En UMAP Workshops.

HANNAK, A., et al. *Measuring personalization of web search*. En Proceedings of the 22nd international conference on World Wide Web. ACM, 2013, p. 527-538.

JOHNSON, M. S. *Personalized Recommendation System for Custom Google Search*. International Journal of Computer & Mathematical Sciences, 2016, vol 5.

KUMAR, R., et al. *Personalized Web Search Using Browsing History And Domain Knowledge*. IEEE Transactions on Knowledge and Data Engineering. 2014, p. 493-497.

LI, Y., JIEBO, L., MEI, T. *Personalized image recommendation for web search engine users*. Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE, 2014, p. 1-6.

LOVARAJU, D., DEVI, L. *An Ontology Like Model for gathering Personalized Web Information*. International Journal of Advanced Research in Computer Science, 2017, vol. 8, no 3.

MAKVANA, K., SHAH, P., et al. *A novel approach to personalize web search through user profiling and query reformulation*. Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on. IEEE, 2014, p. 1-10.

MAKWANA, K. PATEL, J., et al. *An Ontology Based Recommender System to Mitigate the Cold Start Problem in Personalized Web Search*. International Conference on Information and Communication Technology for Intelligent Systems. Springer, Cham, p. 2017, 120-127. doi: 10.1007/978-3-319-63673-3_15

MATTHIJS, N. R., FILIP. *Personalizing web search using long term browsing history*. Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011, p. 25-34.

MOR, N., et al. *Bloom Cookies: Web Search Personalization without User Tracking*. En NDSS, 2015.

PREETHA, S. S., KN Vimal. *Personalized search engines on mining user preferences using clickthrough data*. Information Communication and Embedded Systems (ICICES), 2014 International Conference on. IEEE, 2014, p. 1-6.

PHAM, Xuan Hau; JUNG, Jason J.; NGUYEN, Ngoc Thanh. *Lexical Matching-Based Approach for Multilingual Movie Recommendation Systems*. Recent Developments in Intelligent Information and Database Systems. Springer, Cham, 2016. p. 149-158.

RAJMANE, A. B. P., Pradeep M.; KULKARNI, Prakash J. *Personalization of Web Search Using Web Page Clustering Technique*. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2016, vol. 9, no 10, p. 2295-2300.

RAKESH, G. H. M., Mrs Pratibha. *Customized Personalized Web Search With Privacy Protection*. International Research Journal of Engineering and Technology (IRJET), 2016, vol. 03.

ROGUSHINA, J. *Means of the semantic search personification on base of ontological approach*. International Journal of Mathematical Sciences and Computing (IJMSC). IJ. Mathematical Sciences and Computing, 2016, vol. 2, no 3, p. 1-20. doi: 10.5815/ijmsc.2016.03.01

ROPHIE, A. S. A., A. *User Preferences Based Personalized Search Engine*. International Journal of Research in Computer Applications and Robotics, 2016, vol.4, p. 6-10.

SEKHARBABU, B., et al. *Profile based personalized web search using greedy algorithms*. ARPN Journal of Engineering and Applied Sciences, 2016, vol. 11, no 9.

SHAFIQ, O. A., REDA; ROKNE, JOHN G. *On personalizing Web search using social network analysis*. Information Sciences, 2015, vol. 314, p. 55-76. doi: 10.1016/j.ins.2015.02.029

SHAPIRA, B. Z., BOAZ. *Personalized search: Integrating collaboration and social networks*. Journal of the Association for Information Science and Technology, 2011, vol. 62, no 1, p. 146-160. doi: 10.1002/asi.21446

SHARDA, D. D., PUNAM. *Design of Research Buddy: Personalized Research Paper Recommendation System*. International Conference on Recents Advanced in Engineering Science and Management, 2017.

SHARMA, S. R., Vijay. *Web Personalization through Semantic Annotation System*. Advances in Computational Sciences and Technology, 2017, vol. 10, no 6, p. 1683-1690.

SHOU, L., et al. *Supporting privacy protection in personalized web search*. IEEE Transactions on Knowledge and Data Engineering, 2014, vol. 26, no 2, p. 453-467.

SINGH, A., et al. *Web Semantics for Personalized Information Retrieval*. United States of America. 2017.

THANGARAJ, M. M. *Agent Based Personalized Semantic Web Information Retrieval System*. *International Journal of Advanced Computer Science and Applications*, 2014, vol. 5, No. 8.

USTINOVSKIY, Y. S., Pavel. *Personalization of web-search using short-term browsing context*. En Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013, p. 1979-1988. doi: 10.1145/2505515.2505679

VANITHA, S. *A personalized web search based on user profile and user clicks*. *International Journal of Latest Research in Science and Technology*, 2013, vol. 2, no 5, p. 78-82.

VERMA, D. K., Barjesh. *Multi Agent Architecture for Search Engine*. *International Journal Of Advanced Computer Science And Applications*, 2016, vol. 7, no 3, p. 224-229.

WANG, M., et al. *A personalized result merging method for metasearch engine*. Proceedings of the 6th International Conference on Software and Computer Applications. ACM, 2017, p. 203-207. doi: 10.1145/3056662.3056698

YANG, H. S., Ian. *Privacy-preserving IR 2015: When information retrieval meets privacy and security*. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015, p. 1157-1158.

ZHANG, F. *Smart Image Search System Using Personalized Semantic Search Method*. ProQuest Dissertations Publishing, 2017.

ZHOU, D., et al. *Enhanced Personalized Search using Social Data*. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, p. 700-710.

Zhou, L. *Personalized Web Search*. Proceedings WWV arXiv:1502.01057v1 [cs.IR], 2015.