

Sistema de Recomendación para la Búsqueda Personalizada en un Repositorio de Trabajos de Titulación

Rodrigo Vences Nava, Víctor Hugo Menéndez Domínguez, Alfredo Zapata González

Universidad Autónoma de Yucatán
Calle 60 No. 491-A x 57 Centro Histórico CP 97000, Mérida, México
{vnavas, mdoming, zgonzal}@uady.mx

Resumen — La producción académica de estudiantes en una institución educativa reúne resultados de investigaciones actuales y originales de diversas temáticas. Hoy en día podemos encontrar repositorios de tesis digitales en Internet disponibles para su consulta. Se han hecho esfuerzos por estandarizar dichos repositorios para establecer interoperabilidad entre diferentes instituciones; Sin embargo estos esfuerzos se centran más en la recuperación de información que en la recomendación de contenido. El modelo que se propone es una solución para la búsqueda, recuperación y recomendación de trabajos de titulación en una institución educativa, personalizando la recuperación de información utilizando filtrados por contenido y colaborativo. El sistema de recomendación resultante, producto de esta investigación, se espera facilite el acceso de forma particular a los trabajos de titulación del área de ciencias sociales tanto a profesores como estudiantes permitiendo realizar consultas sobre temas específicos, recuperando la información solicitada y personalizándola de acuerdo al perfil de usuario construido con base en su interacción con el sistema.

Palabras clave — Sistema de Recomendación, Repositorios de Tesis Electrónicas, Bibliotecas Digitales.

I. JUSTIFICACIÓN DE LA PROPUESTA

El crecimiento exponencial que ha tenido Internet [1] en la última década junto con la creciente disponibilidad de documentos en formato digital, almacenados muchas veces de manera no estructurada, ha planteado un gran desafío científico y tecnológico en muchos aspectos, en particular para manipular, buscar y recuperar la información y el conocimiento contenido [2].

Las tesis generadas por las Instituciones de Educación Superior (IES) mexicanas constituyen uno de los acervos más grandes y en constante crecimiento [3]. Sin embargo dicho recurso que ha estado ahí, listo para ser reglamentado, normalizado, almacenado, preservado y organizado para su pronta localización, acceso y uso frecuente en redes distribuidas, ha sido desaprovechado, tanto por su potencial innovador, como por su capacidad para dar visibilidad a egresados, investigadores e instituciones [4].

Esta propuesta se enfoca al área de conocimiento de las ciencias sociales. De forma particular las ciencias antropológicas. Es por ello que la experimentación se realizó en la Facultad de Ciencias Antropológicas de la Universidad Autónoma de Yucatán, la cual cuenta con cinco programas de licenciaturas (Antropología Social, Arqueología, Historia, Comunicación Social, Literatura Latinoamericana y Turismo), y un programa de doctorado en Ciencias Antropológicas [5], los cuales tienen como requisito de egreso la entrega de un trabajo de titulación para obtener el grado de estudios de la disciplina que estén cursando, por ejemplo, tesis, monografía,

proyecto integrador, etc. Estos trabajos se almacenan como una copia en formato físico en la biblioteca del campus de ciencias sociales económico-administrativas y humanidades de la Universidad Autónoma de Yucatán (UADY) y en años recientes en formato digital en las coordinaciones de cada uno de los programas de estudio.

La Facultad de Ciencias Antropológicas de la Universidad Autónoma de Yucatán mantiene una base de datos de alumnos egresados junto con la información de los trabajos de titulación presentados por los mismos desde octubre del año 2007 [5].

Cabe destacar que la UADY dispone de un sistema bibliotecario en línea [6], el cual incorpora búsquedas de tesis, sin embargo esta búsqueda no se realiza basándose en el contenido de los documentos de titulación y tampoco ofrece recomendaciones de trabajos similares. En la Fig. 1 se presenta una captura de pantalla del sistema bibliotecario de la UADY. El usuario puede buscar un trabajo de titulación a partir de ciertos metadatos utilizados para describir un documento (título, autor, tema, serie)

En este trabajo se presenta una propuesta para la búsqueda y recuperación de trabajos de titulación, empleando un sistema de recomendación híbrido que tome diversas fuentes de datos para emitir sugerencias de trabajos de titulación relacionados. Se pretende utilizar toda la información disponible sobre los trabajos de titulación y los usuarios del sistema.

La utilización de toda esta información permitiría mejorar y personalizar las búsquedas, de forma que se puedan ofrecer recomendaciones basadas no sólo en trabajos de titulación de contenidos similares, sino también en los trabajos más utilizados, o en los mejor valorados.

El presente trabajo forma parte de la propuesta de Tesis para obtener el grado de Maestro en Ciencias de la Computación por parte de la Facultad de Matemáticas de la Universidad Autónoma de Yucatán.

El contenido del artículo está organizado de la siguiente manera. En la sección II se presenta el estado actual del conocimiento sobre el tema. En la sección III se detalla el entorno de la propuesta, explicando cada uno de sus componentes, sus relaciones y el flujo de datos entre ellos. En la sección IV se describe la implementación del prototipo. Posteriormente en la sección V se explica la experimentación y se muestra un análisis de los resultados obtenidos. Finalmente la sección VI aborda las conclusiones y las sugerencias para trabajos futuros.

II. ESTADO ACTUAL DEL CONOCIMIENTO SOBRE EL TEMA

En la actualidad hay varias universidades en el mundo que desarrollan colecciones digitales. Por ejemplo, *Networked Digital Library of Theses and Dissertations* (NDLTD) es una

iniciativa para la creación de una biblioteca digital de tesis y disertaciones en la que participan más de 160 universidades alrededor del mundo [7].

No.	Año	Título	Autor	Inf.
1	[1976?]	F 1435.3 .A8 .B67 Horóscopos mayas, o, El pronóstico de los 20 signos del Tzolkin, según los libros de Chilam Balam, de Kaua y de Mani/ Alfredo Barrera Vázquez..	Barrera Vázquez, Alfredo.	Libros (L)
2	c1979.	F 1435.3 .P6 .C63 Codex Perez and the book of Chilam Balam of Mani/ translated and edited by Eugene R. Craïne and Reginald C. Reindorp..		Libros (L)
3	1974.	F 1435.3 .B3 .A48 Textos coloniales del Libro de Chilam Balam de Chumayel y textos glíficos del códice de Dresde/ María Cristina Alvarez Lomeli..		Libros (L)
4	1981.	HB 1992 .Y8 .B34 La migración en el área de los centros coordinadores del IHI de Yucatán : el bracerismo regional y sus repercusiones sociales/ Gilberto Balam ..	Balam Pereira, Gilberto.	Libros (L)
5	1983.	SK 301 .Y33 La cacería del venado : (bilingüe : maya-español)/ David Yah Balam ..	Yah Balam, David.	Libros (L)
6	2003.	Diseño de un paquete didáctico de introducción en la tutoría / María de la Cruz Moreno Balam .	Moreno Balam, María de la Cruz.	Tesis (T)
7	1991.	LE 7.53 .P7k .A57 Antecedentes y perspectivas de la Sección de Ecología Humana del CINVESTAV-U. Mérida/ Raúl Murguía ... [et al.] : comp. por Gilberto Balam ..		Libros (L)

Fig 1. Sistema bibliotecario en línea de la UADY.

La iniciativa denominada La Referencia [8], es una red federada de repositorios institucionales de publicaciones científicas que abarca más de 200 mil tesis de maestría, más de 90 mil tesis de doctorado y cerca de 500 mil artículos, donde participan nueve países, nueve redes nacionales y casi un centenar de instituciones.

En México tenemos la Red Mexicana de Repositorios Institucionales (REMEDI) [9], que tiene por objeto crear una red interconectada de repositorios digitales de Instituciones de Educación Superior (IES) en México, con la finalidad de integrar, difundir, preservar y dar visibilidad a su producción científica, académica y documental, así como también, incorporarse a redes o directorios de repositorios internacionales para fomentar la colaboración y apoyar el acceso y la divulgación de contenidos de acceso abierto.

Dentro del entorno de la UADY, la Facultad de Educación ha implementado un repositorio que aglutina todos los trabajos de titulación que han generado los programas académicos de posgrado que se imparten en la misma. Esta iniciativa es la primera en la institución con estas características [10].

La iniciativa que plantea esta propuesta, es la conformación de un repositorio específico para la Facultad de Ciencias Antropológicas de la misma institución y adicionalmente implementa un grado de personalización en la búsqueda y recuperación de los trabajos de titulación.

En cuanto a los conceptos relacionados con la propuesta, se mencionan los siguientes:

La recuperación de información (RI) significa encontrar material (usualmente documentos) de una naturaleza no estructurada (usualmente texto) que satisface una necesidad de información desde dentro de grandes colecciones (generalmente almacenados en computadoras) [11].

Dentro de los esfuerzos que se han realizado para gestionar la información textual se encuentran principalmente los del área de la RI, la cual trata con la representación, almacenamiento, organización y acceso de elementos de información. Los documentos recuperados deben satisfacer las necesidades de información del usuario expresadas en lenguaje natural [12].

El Procesamiento de Lenguaje Natural (PLN) juega un papel primordial en la mejora de la eficacia en el uso de

herramientas disponibles para el acceso y la búsqueda de información en la Web (por ejemplo usando sinónimos de palabras contenidas en una petición de búsqueda) aprovechando así sus aspectos semánticos [2]. El enfoque principal de esta área consiste en crear métodos, técnicas y herramientas computacionales que permitan realizar análisis de información escrita u oral y que faciliten la búsqueda y organización de dicha información [13].

Algunas de las tareas involucradas en este pre procesamiento lingüístico son [14]:

1. *Transformación previa.* Se refiere a la extracción de la información almacenada en el documento considerando su estructura y formato, de manera que pueda ser procesada más fácilmente.
 2. *Normalización de la información.* A partir de la información previa se realiza un análisis léxico para convertir las cadenas de caracteres en cadenas de palabras (*tokens*) separadas por elementos predefinidos (espacios en blanco, tabuladores, saltos de línea) [15].
 3. *Detección del idioma.* Muchas de las operaciones de procesamiento de información requieren reconocer el idioma de los distintos elementos de un documento estructurado.
 4. *Corrección ortográfica.* Su propósito es verificar que todas las palabras utilizadas correspondan al idioma en el que están expresadas.
 5. *Reducción del vocabulario.* Esta actividad implica eliminar las palabras vacías, extraer la raíz léxica de las palabras y la selección de las palabras clave. Las palabras vacías (*stop words*) son aquellas que no aportan significado al documento sino que se utilizan para seguir las reglas del idioma. La reducción de palabras a su raíz léxica se realiza eliminando prefijos y sufijos mediante un proceso denominado lematización o *stemming* [16]. La selección de las palabras clave consiste en tomar aquellas palabras que son lo suficientemente relevantes para identificar a un documento sin que por ello se reduzca de forma notable el número de documentos que pueden ser recuperados utilizando esas mismas palabras.
- Sin embargo, para un sistema de recuperación de información (SRI), el documento recuperado puede no adaptarse de forma exacta a las peticiones de búsqueda. El SRI debe “interpretar” el contenido de los documentos y clasificarlos de acuerdo al grado de relevancia (Fig. 2). Esta “interpretación” involucra la extracción de información sintáctica y semántica del documento y el cómo utilizar esta información para decidir su relevancia.
- Un sistema de recomendación (SR), recopila información sobre las preferencias de los usuarios sobre un conjunto de elementos (películas, canciones, libros, etc.), desde diferentes fuentes de información para proporcionar a los usuarios predicciones y recomendaciones de elementos similares, tratando de equilibrar factores como exactitud, novedad, dispersión y estabilidad en las recomendaciones [17].
- Existen dos arquitecturas básicas para los SR: por una parte están los sistemas basados en contenido que se enfocan en las propiedades de los ítems. La aproximación de los ítems es determinada mediante la medida de similitud de sus propiedades; por otro lado tenemos los sistemas con filtrado colaborativo que se enfocan en las relaciones entre usuarios e ítems. La aproximación de los ítems es determinada por la similitud de las valoraciones de esos ítems por los usuarios que los han utilizado [18,19].

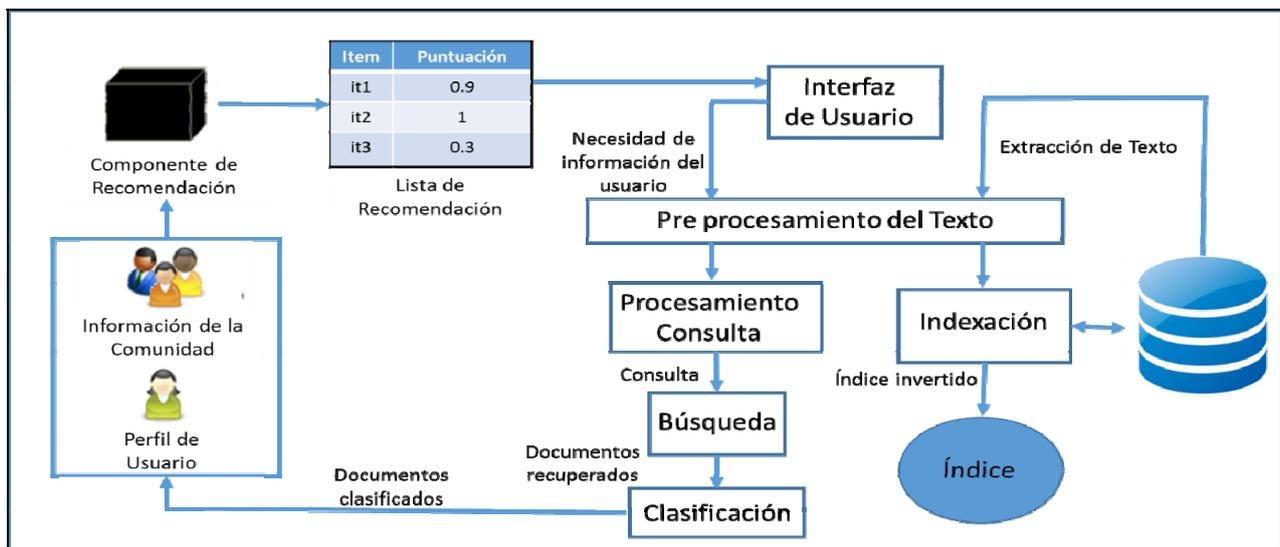


Fig. 2. Proceso de recuperación y recomendación de información. Adaptado de [12].

Numerosos investigadores han adoptado enfoques híbridos entre los SR basados en contenidos y colaborativos, para reducir las desventajas y aprovechar las ventajas que cada uno presenta por separado [17,19, 20].

Un aspecto a tener en cuenta en los SR, es acerca de la obtención de la información, la cuál puede ser de forma explícita o implícita. Un enfoque explícito interacciona directamente con los usuarios a través de un proceso de retroalimentación en el que especifica lo que desean, por ejemplo interrogando a los usuarios mediante un formulario. Un enfoque implícito realiza inferencias a partir de algún tipo de observación sobre los usuarios, por ejemplo analizando el comportamiento de los mismos [21].

En esta propuesta se plantea la creación de un sistema de recomendación híbrido que utilice tanto la información contenida en los trabajos de titulación, así como sus metadatos y la interacción de los usuarios, para recomendar trabajos de titulación del área de ciencias antropológicas que mejor se adapten a las necesidades de información del usuario.

El sistema de recomendación híbrido propuesto utilizará un filtrado basado en contenido y un filtrado colaborativo para atenuar las desventajas que tiene cada filtrado si se utiliza por separado [19]. De esta manera si no contamos con suficiente información de un usuario acerca de sus preferencias podemos utilizar las valoraciones de la comunidad de usuarios para recomendarle los trabajos de titulación más usados o los mejor valorados. Por otro lado si un trabajo de titulación no ha sido usado o valorado, puede ser recomendado de acuerdo a la similitud de su contenido con la búsqueda del usuario.

III. ENTORNO DE LA PROPUESTA

En esta sección se describe el objetivo general y los específicos, así como también, se describen las características generales de la propuesta y las primeras fases de su implementación.

A. Objetivos

El objetivo principal de esta propuesta es desarrollar un sistema de recuperación de trabajos de titulación. Adicionalmente, el sistema recomendará aquellos trabajos de titulación relacionados con los que ha visualizado anteriormente, así como también, se recomendarán los más visualizados por todos los usuarios en conjunto.

En particular, los objetivos que pretenden cumplir son:

- Desarrollar un motor de búsqueda de trabajos de titulación basado en el contenido de documentos.
- Vincular el motor de búsqueda con los metadatos asociados de los trabajos de titulación almacenados en la base de datos.
- Seleccionar una población de profesores y alumnos para realizar pruebas de desempeño del sistema.
- Ejecutar una serie de experimentos que permitan validar la propuesta en términos de precisión y exhaustividad.
- Utilizar un instrumento que mida la usabilidad del sistema y analizar los datos obtenidos de la interacción de los usuarios con el sistema.

B. Características generales

En la Fig. 3 se presenta la arquitectura de la propuesta. Se pretende probar que la búsqueda, selección y recuperación de trabajos de titulación puede ser más eficiente si se realiza a través de un modelo híbrido de recomendación, cuyas recomendaciones se apoyen en diversas fuentes de información asociadas a los metadatos y el contenido de los trabajos de titulación, así como también, las interacciones que han realizado los usuarios con el sistema.

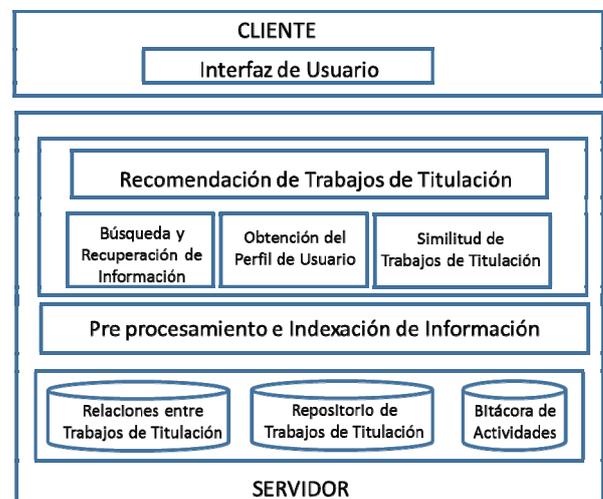


Fig. 3. Representación de la arquitectura propuesta.

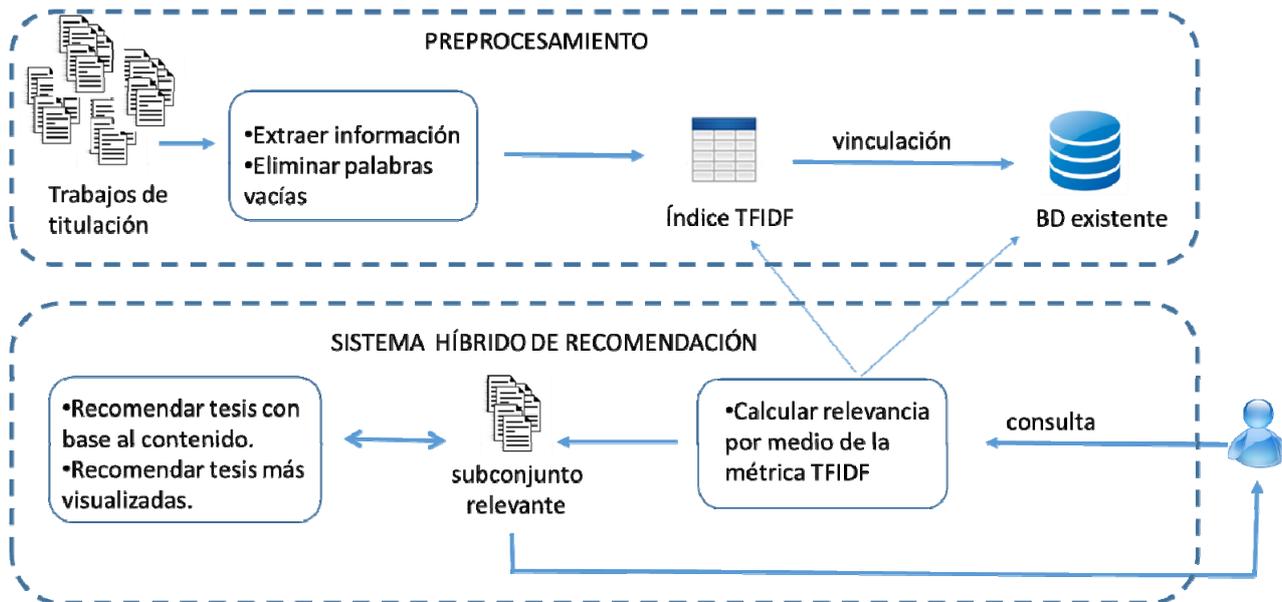


Fig. 4. Esquema del funcionamiento del modelo.

En la arquitectura podemos observar el repositorio de trabajos de titulación sobre el cuál se realizará el pre procesamiento para obtener la representación vectorial de los documentos, con la cual se realizarán los cálculos de similitud entre ellos. También se observa la bitácora de actividades que almacena toda la interacción de los usuarios con el sistema y servirá para crear los perfiles de los usuarios. En la parte más alta de la arquitectura se encuentra el sistema de recomendación propiamente, que es el encargado de decidir la relevancia de los documentos obtenidos del sistema de recuperación, con base en el perfil de usuario, la valoración y el uso de los trabajos de titulación.

Cuando el usuario realice consultas al sistema de recomendación a través de palabras clave, se comunicará con el módulo de búsqueda y éste devolverá un listado de trabajos de titulación ordenados de acuerdo al grado de similitud entre los términos más representativos de los documentos de titulación y la consulta del usuario. Los términos más representativos se calcularán utilizando el modelo vectorial [22].

La Fig. 4 muestra el esquema del funcionamiento del modelo, para el cuál primero es necesario realizar un pre procesamiento para después generar la recomendación.

Como parte del pre procesamiento de la información, se creará un índice invertido de términos con base al contenido de los trabajos de titulación recuperados, los cuales provienen de los programas de licenciatura de la Facultad de Ciencias Antropológicas de la UADY [5]. La tarea de construcción del índice involucra un pre procesamiento de la información [23]. Una vez creado el índice de términos, se vinculará a la base de datos para recuperar la información que el usuario solicite, considerando el contenido de los trabajos de titulación y sus metadatos asociados (autor, título del trabajo, licenciatura, año).

También, contará con un módulo que determine la similitud entre los trabajos de titulación para enriquecer la predicción y recomendación de los mismos, este proceso lo hará con respecto a su contenido, utilizando como medida de similitud la métrica TF-IDF, basada en la función *match* de Mysql [24], la cual utiliza una medida de relevancia basada en la ocurrencia de los términos dentro del documento por medio de la Eq. 1:

$$w = tf * idf. \quad (1)$$

Donde:

- w = es el peso del término
- tf = es la frecuencia del término en el documento
- idf = es la frecuencia del término dentro de la colección de documentos

Adicionalmente, requerirá de un módulo para obtener la información del perfil del usuario, es decir, las preferencias del usuario con base en el registro de actividad con el sistema. El sistema de recomendación obtendrá la información de los usuarios de manera implícita mediante el registro de las acciones que se producen entre la interacción del profesor o estudiante y el entorno, identificando a cada uno por medio de su dirección IP y el uso de cookies como elemento de registro e identificación que resulta transparente para el usuario.

El sistema de recomendación utilizará todos los módulos antes descritos para generar predicciones de trabajos de titulación cuando el usuario se autentique en el sistema, tomando en cuenta el historial de su interacción con el sistema. En caso de que no cuente con información histórica utilizará la información de la comunidad con relación a los trabajos más usados o los más valorados. También, generará recomendaciones de trabajos de titulación conforme el usuario use o valore algún trabajo, utilizando la información de la similitud de trabajos de titulación. Con respecto al uso de un trabajo de titulación, se refiere a la visualización del documento en el navegador web. En cuanto a la valoración del usuario, se considerará una escala de aprobación similar a lo que plantean las redes sociales, asignando únicamente valoraciones positivas del estilo "me gusta".

Para calcular la similitud entre los documentos se obtendrá el término más representativo de los últimos tres documentos visualizados y con base a ese término se generará una búsqueda al sistema. Adicionalmente, se utilizarán los cinco primeros resultados de cada búsqueda para ser propuestos inicialmente como recomendación de trabajos similares. El usuario podrá consultar un listado mayor si es de su interés.

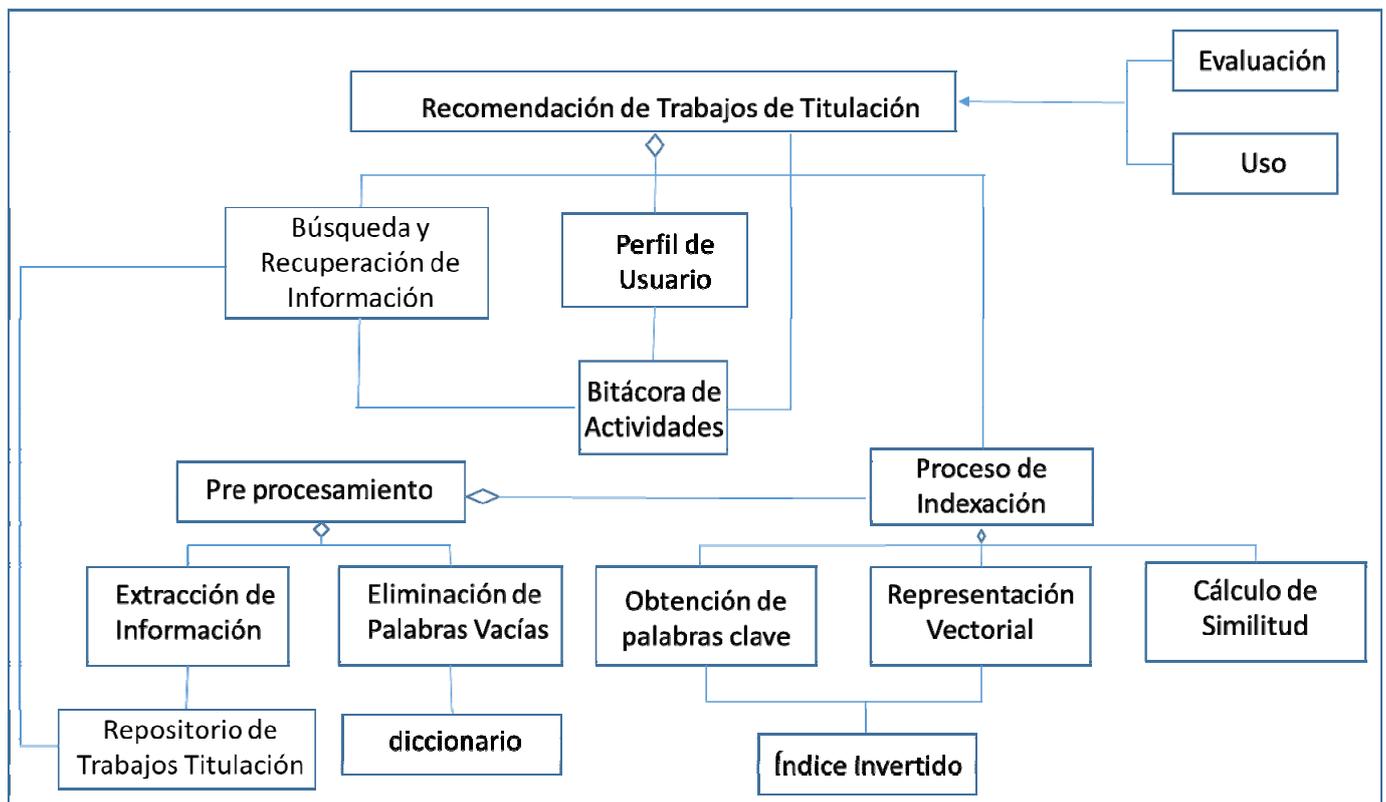


Fig. 5. Diagrama de clases del sistema de recomendación

El término más representativo de un documento es aquél que tenga la mayor frecuencia de aparición dentro del mismo.

Para la recomendación de los trabajos de titulación más usados, se hará uso de la información de registro de todos los usuarios, presentando una lista de aquellos trabajos que los profesores y alumnos hayan visualizado en más ocasiones por medio de su navegador.

La visualización de los trabajos de titulación se contará una única vez independientemente de las veces que lo haya consultado el usuario en el navegador web.

Es importante destacar que el módulo de recomendación permitirá al usuario realizar la modificación del orden de los trabajos de titulación presentados de acuerdo a los criterios: similitud de contenido, uso y evaluación. Existirá un módulo gestor de trabajos de titulación que se encargará de administrar el ingreso y/o modificación de trabajos de titulación a la base de datos. Este último módulo descrito y el de autenticación son externos a la propuesta pero necesarios para su funcionamiento, por lo cual se describen pero no aparecen en la Fig. 5 que muestra el diagrama de clases de la propuesta.

Para el almacenamiento de los metadatos se ha optado por el estándar Dublin Core [25], ya que permite la interoperabilidad entre repositorios y se enfoca a la descripción de cualquier documento digital. Su uso es muy extendido actualmente.

El estándar Dublin Core (DC) define un conjunto base de 15 metadatos que constituyen el nivel simple. Estos metadatos describen el contenido del documento, la propiedad intelectual y la información para la creación de instancias.

El nivel cualificado incorpora 3 nuevos metadatos (audiencia, provenance y rightHolder) al nivel simple. También define un conjunto de atributos que clarifican la representación semántica de los recursos descritos para una

fácil recuperación. Estas propiedades refinan las descripciones de metadatos y están limitadas a ciertos elementos, pudiendo utilizar vocabularios asociados o descripciones en lenguaje natural.

En DC, cada elemento es opcional y puede repetirse. No hay un orden específico de aparición. Su contenido puede regirse por esquemas de codificación, vocabularios controlados o estar abierto a las descripciones de texto en lenguaje natural.

El estándar proporciona un conjunto de directrices para su implementación en los formatos XHTML, XML y RDF, lo que permite la interoperabilidad de los recursos [26].

IV. IMPLEMENTACIÓN

Para la implementación del modelo se utiliza el lenguaje de programación PHP y el manejador de base de datos MySQL en un servidor Apache. Todo el software es de libre distribución y multiplataforma [27].

La página principal del sistema (Fig. 6) muestra la versión preliminar de la recomendación colaborativa de los trabajos de titulación más visualizados.

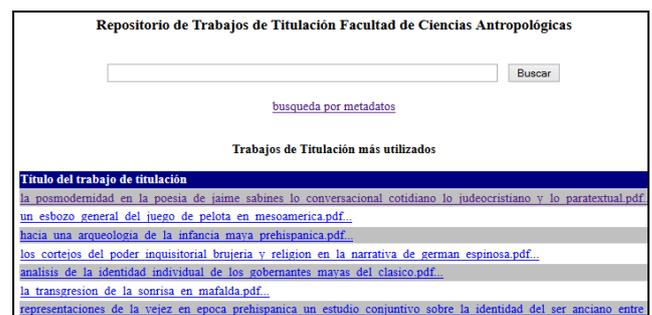


Fig. 6. Pantalla principal del sistema

Cuando un usuario realiza una consulta, se despliega una lista de los trabajos de titulación recuperados de la base de

datos, los cuales están ordenados por relevancia de los términos de la cadena de búsqueda (Fig.7, lado izquierdo). Adicionalmente, se muestra las recomendaciones de los trabajos de titulación con base en los tres últimos documentos visualizados (Fig. 7, lado derecho)

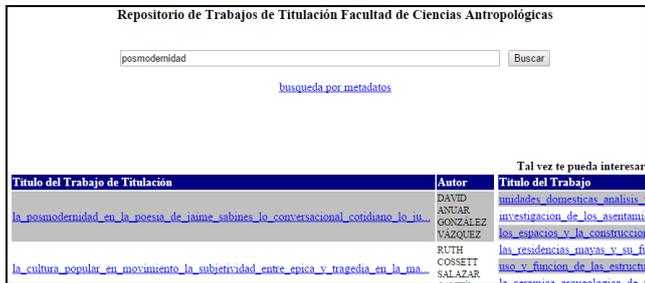


Fig. 7. Ejemplo de recomendación basada en contenido.

V. EXPERIMENTACIÓN

En este apartado se presenta una prueba inicial realizada para valorar la usabilidad del sistema de recomendación desarrollado. En primera instancia se presenta el objetivo, seguidamente se explica cómo fue desarrollado el caso de estudio y finalmente se dan a conocer las limitaciones y los resultados obtenidos.

A. Objetivo

El objetivo que se deseaba analizar en este caso de estudio ha sido: evaluar la usabilidad del sistema de recomendación y por ende el modelo propuesto.

Por usabilidad nos referimos a un atributo de calidad que establece cuán fácil y útil resulta utilizar un sistema [28]. Esto, implica valorar si los usuarios interactúan con el sistema de la forma más fácil, cómoda e intuitiva posible [29].

B. Diseño del estudio

Se realizó un estudio donde los sujetos tenían que utilizar el sistema de recomendación realizando consultas a cerca de temas de su interés con el fin de comprobar si el sistema recuperaba trabajos de titulación relevantes a la consulta y a su vez recomendará trabajos similares al consultado en ese momento o en anteriores consultas.

C. Sujetos

El grupo interdisciplinario de participantes estuvo integrado por 17 profesores de la Facultad de Ciencias Antropológicas de la UADY. Sus edades fluctúan entre los 30 y 40 años y pertenecen a las áreas de antropología social (5), arqueología (4), historia (2), comunicación social (3) y literatura latinoamericana (3). La mayoría de ellos tiene estudios de posgrado (16) y doctorado (6) en su área de conocimiento.

D. Procedimiento

Los sujetos fueron visitados en su espacio de trabajo y se les indicó la URL del sistema a la que deberían ingresar desde su computadora y se les explicó de manera breve el funcionamiento del mismo y la forma de utilizarlo. Se les solicitó que realizaran consultas sobre temas de su interés y seleccionaran (visualizaran en el navegador) los trabajos de titulación que llamaran su atención.

Al finalizar las consultas, los participantes respondieron un cuestionario que valida la usabilidad del sistema.

E. Resultados

Para el registro de la percepción de los usuarios en términos de la usabilidad de un sistema, se empleó un cuestionario estándar *System Usability Scale* (SUS) [30]. Este instrumento consta de 10 reactivos como se muestra en la tabla 1. Cada elemento emplea una escala de Likert de 5 puntos en un rango de 1 (completamente en desacuerdo) hasta 5 (completamente de acuerdo).

Encuesta SUS	Promedio
1. Volveré a usar esta aplicación.	3.9
2. La aplicación es compleja.	3.6
3. La aplicación es fácil de usar.	3.9
4. Es necesario el apoyo de un experto para usar la aplicación.	3.6
5. Las opciones de la aplicación están bien integradas.	3.5
6. La aplicación es inconsistente.	3.7
7. Aprendí rápidamente a usar la aplicación.	3.8
8. Es incómodo usar la aplicación.	3.7
9. Me gustó usar la aplicación.	3.8
10. Necesito aprender muchas cosas antes de usar la aplicación.	3.7
Usabilidad	92.86%

Tabla 1. Resultado de la encuesta SUS.

La propia prueba SUS tiene su procedimiento para calcular su resultado en un rango de 0 a 100%. Un resultado arriba de 70% se puede considerarse como aceptable.

El resultado del cuestionario (14 de 17 profesores evaluaron) nos proporciona como resultado una usabilidad del 92.86% que está muy por encima del mínimo considerado como aceptable.

Por otra parte, a partir de la información registrada por el sistema en cuanto a la interacción de los usuarios se hizo un gráfico que se presenta en la Fig. 8, la cual muestra la relación entre usuarios y trabajos de titulación descargados. Los cuadrados son los usuarios y los círculos son los trabajos de titulación. El área de los círculos se basa en la cantidad de descargas de las tesis. Como un caso de estudio, el triángulo seleccionado es el trabajo de titulación más visualizado (4 veces) y está conectado a los 4 usuarios distintos (direcciones IP) que lo visualizaron.

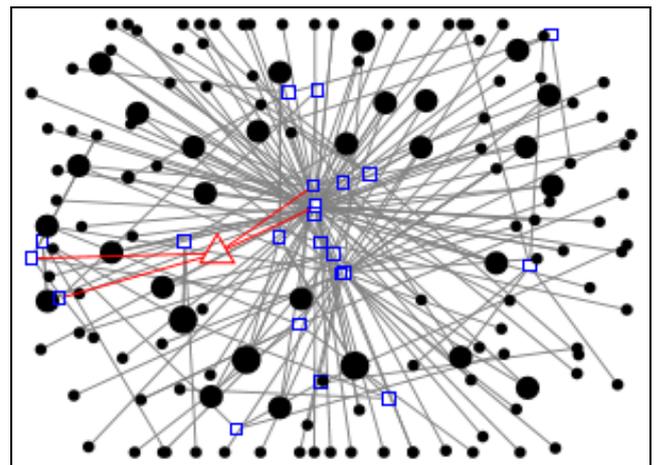


Fig. 8. Relación entre usuarios y tesis descargadas.

Para realizar un análisis más profundo se plantea como trabajo futuro realizar pruebas de desempeño para comparar los resultados del sistema propuesto con los resultados de los sistemas similares mencionados. De la misma manera se pretende utilizar una encuesta basada en el modelo de Aceptación de Tecnología (TAM) [31], para registrar la percepción de los usuarios con respecto a la utilidad y

facilidad de uso, para complementar y contrastar los resultados de la prueba SUS.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

La consecución de este proyecto se espera que repercuta en la utilización de esta herramienta por parte de los profesores como de los alumnos de la institución para la búsqueda de información de trabajos de titulación existentes referentes a su área de interés.

El sistema que presenta este trabajo ofrece una solución relevante al acceso a la información de trabajos de titulación de la Facultad de Ciencias Antropológicas de la UADY recomendando a los usuarios trabajos de titulación similares, utilizando como métrica de similitud entre documentos la frecuencia de los términos contenidos en el mismo. Por otro lado, el sistema recomienda a los usuarios los trabajos de titulación más visualizados. De este modo disminuye la sobrecarga de información permitiendo al usuario encontrar trabajos de titulación similares a los que solicita en una consulta al sistema.

Una mejora que podría incorporarse al sistema de búsqueda sería que realice expansión de consulta considerando los sinónimos de los términos proporcionados por el usuario.

Como trabajo futuro se proponen varias líneas de desarrollo. Por una parte, se pretende utilizar la información de los perfiles de los usuarios para realizar recomendaciones colaborativas basadas en los mismos. También, se pretende encontrar mejores mecanismos de identificar a los usuarios en combinación con la dirección IP.

Por otra parte, se implementarán técnicas de minería de datos con el objetivo de identificar grupos basados en los trabajos de titulación consultados. Adicionalmente, se pretende extraer reglas de asociación y de clasificación con el objetivo de recomendar la consulta de trabajos similares que puedan ser de interés.

En cuanto a la interfaz del sistema propuesto, se realizarán mejoras de manera que una sola página permita realizar tanto la búsqueda por contenido como la búsqueda por metadatos. Para ello se establecerán mecanismos para la adecuada visualización y/o descarga de los trabajos de titulación.

Otro aspecto importante a incorporar en la interfaz, es que se permitirá a los usuarios asignar una valoración a los trabajos de titulación consultados, para realizar recomendaciones colaborativas basadas en estas puntuaciones.

Finalmente, se pretende conectar los repositorios de trabajo de titulación de las Facultades de Educación y de Ciencias Antropológicas. Este sería el inicio de la conformación de un repositorio para el campus de ciencias sociales económico administrativas de la UADY. Por lo tanto, los avances obtenidos en esta investigación se extenderán a otras áreas específicas de conocimiento.

REFERENCIAS

[1] Internet world stats. <http://www.internetworldstats.com/stats.htm>. Accedido el 12 de mayo de 2014.

[2] Olivas, J. Búsqueda eficaz de información en la Web, Editorial de la Universidad Nacional de la Plata (Edulp). Buenos Aires, Argentina. (2011).

[3] Castro Thompson, A.; Alonso Gamboa, J.O.; Zetter Leal, J. Repositorio dinámico colaborativo: red de bibliotecas

digitales – ECOES. XXXIX Jornadas Mexicanas de Biblioteconomía. Universidad Autónoma de Chihuahua. (2008).

[4] Arreola Semadeni, M. H. Repositorios de tesis digitales de posgrado: análisis en trece instituciones mexicanas. (2012). Tesis de Maestría.

[5] Sitio WEB de la Facultad de Ciencias Antropológicas de la Universidad Autónoma de Yucatán, <http://www.antropologia.uady.mx/titulados/index.php>, Accedido el 26 de Noviembre de 2014.

[6] Sistema Bibliotecario Universidad Autónoma de Yucatán, <http://www.bibliotecas.uady.mx>, Accedido el 26 de Noviembre de 2014.

[7] Networked Digital Library of Theses and Dissertations. <http://www.ndltd.org/>. Accedido el 26 de Noviembre de 2014.

[8] Red Federada de Repositorios Institucionales de Publicaciones Científicas. <http://lareferencia.redclara.net/rfri/>. Accedido el 26 de Noviembre de 2014.

[9] Red Mexicana de Repositorios Institucionales. <http://www.remeri.org.mx/>. Accedido el 26 de Noviembre de 2014.

[10] Zapata A.; Menéndez V.; Campos M.; Hernández G.; Flores H. Repositorio de Trabajos de Titulación de la Facultad de Educación de la Universidad Autónoma de Yucatán. *XXI Encuentro Internacional de Educación a Distancia*. Universidad de Guadalajara (2013).

[11] Manning, C., R Raghavan, P. y Schütze H., *An Introduction to Information Retrieval*, Cambridge University Press, Cambridge, England, 2009.

[12] Baeza-Yates, R. A. y Ribeiro-Neto, B., *Modern Information Retrieval*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.

[13] Verspoor K., Cohen K., Natural Language Processing. *Encyclopedia of Systems Biology*, p. 1945 – 1948, (2013).

[14] Romero, F. P. Metodología basada en técnicas de soft-computing para la organización y gestión de documentos de naturaleza heterogénea. Universidad de Castilla-La Mancha, Ciudad Real - España. (2008).

[15] Fox, C. Lexical analysis and stoplists. In W. Frakes, y R. Baeza-Yates (Eds.), *Information retrieval. Data structures & algorithms* (pp. 102-130). New Jersey: Prentice-Hall Inc. (1992).

[16] Frakes, W. Stemming algorithms. In W. Frakes, y R. Baeza-Yates (Eds.), *Information retrieval. Data structures & algorithms* (pp. 131-160). New Jersey: Prentice-Hall Inc. (1992).

[17] Bobadilla J.; Ortega F.; Hernando, A.; Gutiérrez A. Recommender systems survey. Universidad Politécnica de Madrid. Elsevier. (2013).

[18] Leskovec J., Rajaraman A. y Ullman, J., *Mining of Massive Datasets*, Stanford University, 2014, Disponible en <http://www.mmids.org/>.

[19] Ricci F., Rokach L., Shapira B. y Kantor P., *Recommender Systems Handbook*, Springer New York. Dordrecht Heidelberg London, 2011.

[20] Nieves C., Sistema de Recomendación para la Composición de Objetos de Aprendizaje, 2014, Tesis de Maestría.

[21] Porcel C., Herrera-Viedma y Morales J., “Mejorando el sistema de recomendaciones SIRE2IN: un enfoque interdisciplinario”, II SIMPOSIO SOBRE LÓGICA FUZZY Y SOFT COMPUTING, pp. 81-88, 2007.

[22] Salton, G., *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*, Addison Wesley, 1989.

[23] Romero Ch., Metodología Basada en Técnicas de Soft-Computing para la Organización y Gestión de Documentos de

Naturaleza Heterogénea, Universidad de Castilla de la Mancha. Tesis Doctoral, 2008.

- [24] MySQL Internals Manual :: 10 Important Algorithms and Structures :: 10.7 Full-Text Search.: <http://dev.mysql.com/doc/internals/en/full-text-search.html>.
Accedido el 3 de Diciembre de 2014.
- [25] DCMI. Dublin core metadata initiative. (2008). <http://www.dublincore.org>. Accedido el 9 de Junio de 2015.
- [26] DCMI. Dublin core metadata initiative encoding guidelines. (2008). <http://www.dublincore.org/resources/expressions>.
Accedido el 28 de Agosto 2015.
- [27] Lemire D. y McGrath S., “Implementing a Rating-Based Item-to-Item Recommender System in PHP/SQL”, Technical Report D-01, (2005).
- [28] Nielsen, J., y Loranger, H., *Prioritizing web usability*, Berkeley: New Riders Press, 2006.
- [29] Dumas, J., y Redish, J. *A practical guide to usability testing*. Portland: Intellect Books, 1999.
- [30] Brooke, J., “SUS – A quick and dirty usability scale”, http://cui.unige.ch/isi/icle-wiki/_media/ipm:test-suschapt.pdf,
Accedido el 3 de Diciembre de 2014.
- [31] Tang, D., y Chen, L. “A review of the evolution of research on information technology acceptance model”. In *Business Management and Electronic Information (BMEI), 2011. International Conference on, 13-15 Mayo 2011* (Vol. 2, pp. 588 - 591): IEEE Computer Society.



Rodrigo Vences Nava es Licenciado en Ciencias de la Computación por la Universidad Autónoma de Yucatán, México. Actualmente se encuentra cursando la Maestría en Ciencias de la Computación de la Universidad Autónoma de Yucatán, México. Es Administrador de Tecnologías de Información de la Facultad de Ciencias Antropológicas de la Universidad

Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con repositorios de tesis electrónicas, sistemas de recomendación y minería de datos.



Víctor Hugo Menéndez Domínguez es Doctor en Tecnologías Informáticas Avanzadas por la Universidad de Castilla-La Mancha, España. Es Profesor Titular en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en

temas relacionados con la Educación a distancia, la representación del conocimiento y la gestión de Objetos de Aprendizaje.



Alfredo Zapata González es Doctor en Tecnologías Informáticas Avanzadas por la Universidad de Castilla-La Mancha, España. Es Profesor Titular en la Facultad de Educación de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas

relacionados con Minería de datos y Sistemas de Recomendación para la Educación a distancia.