

Recomendación personalizada de documentos de Repositorio Digital Institucional aplicando minería de texto

Díaz, Fátima S., Fernández Reuter, Beatriz y Durán, Elena B.

Instituto de Investigaciones en Informática y Sistemas de Información – Facultad de Ciencias Exactas y Tecnologías – Universidad Nacional de Santiago del Estero (UNSE)

e-mail: fatimadiaz91@gmail.com; bfreuter@unse.edu.ar; eduran@unse.edu.ar

RESUMEN

Para realizar una investigación con material de calidad es fundamental contar con fuentes confiables. Es por ello que muchos recurren a repositorios digitales de universidades, dado que pueden acceder a información corroborada y utilizarla para producir nueva.

A pesar de que muchas universidades tienen sus propios repositorios, no suele sacarse el provecho adecuado de los mismos. Al existir tanta información en un solo lugar, cuesta realizar una exploración exhaustiva haciendo un uso eficiente del tiempo.

Una solución a este problema puede ser la incorporación al repositorio digital, de un sistema de recomendación personalizado, que ayude al usuario a encontrar información de manera más ágil y amigable, fomentando así el uso del repositorio.

Es por esto que se ha iniciado una línea de investigación enfocada en los sistemas de recomendación para repositorios digitales. En particular esta investigación se centra en el desarrollo de un sistema de recomendación de contenido para los usuarios del repositorio digital institucional de la Biblioteca Central de la Universidad Nacional de Santiago del Estero. Se propone la creación de un recomendador personalizado de documentos, basado en contenido, aplicando minería de texto, que contará con un perfil de usuario

para que las recomendaciones puedan ser personalizadas.

Palabras clave

Sistemas de Recomendación, Repositorio Digital Institucional, Personalización, Minería de Texto

CONTEXTO

Esta línea de investigación se desarrolla en el marco del proyecto de investigación “Métodos y técnicas para desarrollo de aplicaciones ubicuas”, código 23/C139, correspondiente a la convocatoria 2017-2020 de la Secretaría de Ciencia y Tecnología de la UNSE (SICYT – UNSE). El mencionado proyecto tiene entre sus objetivos “*Evaluar y proponer técnicas y métodos para el razonamiento y la adaptación de aplicaciones ubicuas*” y “*Aplicar las técnicas, métodos y estrategias propuestos en la construcción de aplicaciones ubicuas*”. Con el propósito de dar cumplimiento a estos objetivos específicos, se identificaron y analizaron problemáticas locales y regionales que potencialmente pueden demandar la construcción de aplicaciones ubicuas. Entre esas problemáticas se identificó la consulta a repositorios digitales desde cualquier lugar y en cualquier momento.

1. INTRODUCCIÓN

Un *Repositorio Institucional* (RI) es un conjunto de servicios Web centralizados, creados para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una institución. (De Giusti, Marisa Raquel, 2013)

En los repositorios institucionales de las universidades se suele buscar información para investigaciones científicas, puesto que la producción académica en dichas instituciones educativas avanzadas reúne resultados de investigaciones de diversas temáticas y es muy prolifera.

En esta nueva era de la digitalización las búsquedas pueden realizarse sin preocuparse por la distancia geográfica y la zona horaria. Haciendo uso de Internet podemos consultar repositorios institucionales de distintas universidades, sin tener la necesidad de trasladarnos a ellas.

Cada vez más universidades cuentan con repositorios propios, depositando en ellos diverso material de investigación, como publicaciones en congresos, revistas, capítulos de libros, tesis, tesinas, etc. A medida que el tiempo avanza, la cantidad de información va aumentando.

Realizar una búsqueda en un repositorio puede llegar a ser cansador, desmotivador y abrumador si no se tiene experiencia en búsqueda en los mismos, o incluso teniendo experiencia suele quedar la duda si la exploración realizada fue suficiente o si queda más material por explorar.

Para brindar solución a esta problemática es que surge la propuesta de prototipo de un recomendador personalizado de documentos para el repositorio digital institucional de la Universidad Nacional de Santiago del Estero.

Un *sistema recomendador* es una tecnología de filtrado de información personalizada, usada para predecir si a un usuario particular le gusta un ítem determinado (problema de predicción), o identificar un conjunto de N ítems que pueden interesarles a ciertos usuarios (problema de recomendación top-N) (Karypis, 2001).

Un sistema de recomendación *basado en contenido* se enfoca en las propiedades de los ítems. Este tipo de técnicas realizan las recomendaciones a partir de la similitud entre el contenido (información y características) de los ítems a recomendar y lo que el usuario ha señalado como de interés, lo cual permitirá personalizar la recomendación.

Existen numerosos trabajos relacionados a los sistemas de recomendación de contenido educativo y material de investigación, entre los que se pueden mencionar a Vences-Nava, Menéndez-Domínguez, & Zapata-González, (2015) quienes proponen un sistema de recomendación personalizado híbrido (combinando el filtrado colaborativo con el filtrado basado por contenido) en un repositorio de trabajos de titulación de la Universidad de Yucatán (México). Dicho sistema utiliza tanto la información contenida en los trabajos de titulación, 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 de los usuarios.

Otro trabajo de sistema de recomendación sobre repositorios digitales, aunque en este caso sobre objetos de aprendizaje, es el realizado por Chacón-Rivas, Mario; Garita, Cesar; Solís, (2015), quienes proponen sugerir OA que podrían ser de utilidad para la lección descrita en un diseño instruccional por un profesor. Dicho sistema de recomendación emplea características propias de los sistemas

colaborativos, demográficos y de los basados en contenidos.

La propuesta de Fernández-Reuter & Durán, (2014), consiste en un buscador automático de material educativo en aulas virtuales que además recomienda, utilizando técnicas de Minería de contenido web, material adicional que tenga relación a una consulta ingresada por un estudiante. En este último trabajo se basará el sistema propuesto.

En base a los trabajos anteriormente mencionados, que forman parte de los cimientos de la propuesta, se pretende demostrar que utilizando un recomendador personalizado basado en contenido y personalizado con etiquetas de interés del usuario, se brindarían mejores sugerencias para las investigaciones realizadas, permitiendo de esta manera que el usuario encuentre los documentos que le pueden resultar de utilidad, logrando resultados más específicos en un menor tiempo.

2. DESCRIPCIÓN DE LA LÍNEA DE INVESTIGACIÓN Y DESARROLLO

En esta investigación se abordan cuestiones relacionadas a la recomendación personalizada de documentos para el Repositorio Digital Institucional de la UNSE.

Para realizar la recomendación se tendrán en cuenta los datos del título, palabras claves y resumen de las publicaciones en congreso, revista, libros, capítulos de libro, tesis, trabajos finales de grado y posgrado que se encuentren almacenados en el repositorio.

Se propone trabajar con técnicas de minería de texto para realizar la recomendación. De esta manera, se entrenará un modelo que sea capaz de reconocer las temáticas asociadas a un documento, lo que se utilizará para conformar un conjunto de documentos a

recomendar al usuario. Para dicho entrenamiento, es necesario contar con la ayuda de un experto que identificará los temas con los que se relaciona cada documento del repositorio.

Se desarrollará un prototipo que permitirá determinar la validez del modelo, dado que la recomendación brindada por el sistema dependerá de la precisión del modelo en la identificación de las temáticas asociadas a la búsqueda.

A su vez, se deberá contar con un perfil de usuario con el fin de personalizar dicha recomendación, el cual contará con los temas de interés del usuario.

Como se trabajará con un repositorio ya existente, se considerarán los lineamientos, metadatos y arquitectura del RDI utilizado, de tal manera que haya una concordancia para una futura implementación.

3. OBJETIVOS Y RESULTADOS ESPERADOS

A pesar de que cada vez es más común que los sistemas acerquen al usuario aquello que puede resultar de su interés, esto no suele aplicarse en los repositorios digitales institucionales. Es por ello que, con este trabajo se espera contribuir a mejorar la difusión, búsqueda y recuperación del material almacenado en el Repositorio Digital Institucional (RDI) de la Biblioteca Central de la UNSE, recomendando automáticamente documentos personalizados según el interés de cada usuario.

Los objetivos específicos de esta investigación son:

- Desarrollar y entrenar un modelo que detecte patrones para clasificar documentos, almacenados en el RDI

de la UNSE, en base al título, resumen y palabras claves del documento.

- Diseñar y desarrollar un prototipo de recomendador personalizado, que haga uso del modelo, para brindar sugerencias personalizadas en base a las etiquetas almacenadas en el perfil del usuario, para ser implementado en el RDI de la UNSE.

Con este trabajo se espera obtener los siguientes resultados:

- Modelo de minería de texto que detecte patrones para clasificar documentos, almacenados en el repositorio digital institucional de la UNSE, en base al título, resumen y palabras claves del mismo.
- Prototipo de recomendador personalizado, que haga uso del modelo, para brindar sugerencias personalizadas en base a las etiquetas almacenadas en el perfil del usuario, para ser implementado en el RDI de la UNSE.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de esta línea de investigación se está formando una Becaria de Grado con Beca de Investigación para Estudiantes, otorgada por la Consejo de Investigación Científica y Tecnológica de la Universidad Nacional de Santiago del Estero (CICyT-UNSE), períodos 2018 y 2019, y se está desarrollando un Trabajo Final de Grado de la carrera Licenciatura en Sistemas de Información de la UNSE.

5. BIBLIOGRAFÍA

Chacón-Rivas, Mario; Garita, Cesar; Solís, J. (2015). *Agente Híbrido Recomendador de Objetos de Aprendizaje. Lacló*, 5. Retrieved from https://www.researchgate.net/publication/270341748_Agente_Hibrido_Recomendador_de_Objetos_de_Aprendizaje

De Giusti, Marisa Raquel (2013). *¿Por qué conviene construir un Repositorio Institucional?*. Reunión del Consejo Interuniversitario Nacional (CIN). Retrieved from <http://sedici.unlp.edu.ar/handle/10915/31682>

Fernández-Reuter, B., & Durán, E. B. (2014). *Buscador automático de material educativo en aulas virtuales*. IX Congreso de Tecnología En Educación y Educación En Tecnología (TE&ET 2014), 166–173. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/38501>

Karypis, G. (2001). *Evaluation of Item-Based Top-N Recommendation Algorithms*, 1–13. Retrieved from <https://dl.acm.org/citation.cfm?id=502627>

Vences-Nava, R., Menéndez-Domínguez, V., & Zapata-González, A. (2015). *Modelo híbrido de recomendación de trabajos de titulación*, (January 2016). Retrieved from https://www.researchgate.net/publication/291333144_Modelo_hibrido_de_recomendacion_de_trabajos_de_titulacion