

Avances en el desarrollo de una herramienta de búsqueda en repositorios académicos

Nélida R. Cáceres, Ana C. Tolaba, Sergio L. Martínez

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 20 San Salvador de Jujuy – 0388 4221576
nrcaceres@fi.unju.edu.ar

RESUMEN

En la actualidad, el Grupo de Investigación y Desarrollo en Ingeniería de Software (GIDIS) lleva a cabo un proyecto de investigación que tiene por objetivo desarrollar una herramienta de búsqueda que facilite el análisis y comprensión de los datos referidos a trabajos finales de grado almacenados en un repositorio digital, el cual posibilitará la difusión de dichos trabajos, ya sea para la comunidad universitaria como así también para la comunidad en general.

El presente proyecto plantea el estudio de repositorios digitales que utilizan las instituciones académicas con el fin de definir todos los conceptos relacionados al mismo y que sean adecuados para el repositorio de la Facultad de Ingeniería. Esto permitirá definir estrategias de búsquedas, técnicas y herramientas que incorporen tecnologías de web semántica y sistemas NoSQL, además, identificar y especificar las características que faciliten la experiencia del usuario durante su interacción con este repositorio. A continuación, se presenta el estado de avance alcanzado en este proyecto, los resultados y la formación de recursos humanos concretada en el marco del mismo.

Palabras clave: *NoSQL, Repositorios Digitales, Semántica, Experiencia del usuario, UX.*

CONTEXTO

La propuesta de desarrollo de una *Herramienta de Búsqueda en repositorios académicos*, se encuentra enmarcada en el proyecto “*Desarrollo de Herramienta utilizando Web Semántica y Sistemas NoSQL*” aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy. Este proyecto es categoría A (código D/0168) y se encuentra bajo incentivo.

1. INTRODUCCIÓN

Un repositorio es un sistema de información que organiza, almacena y brinda el acceso libre a la producción de la información [1]. La tendencia de exponer la producción de las instituciones a través de repositorios digitales de Acceso Abierto [2] ha incrementado la cantidad de repositorios digitales notablemente.

Debido a que los repositorios digitales permiten la creación e intercambio de conocimiento, facilitando la transferencia de conocimiento al sector productivo, su implementación ofrece una variedad de ventajas tanto a investigadores, estudiantes, así como al resto de la sociedad [3].

En el diseño de un repositorio digital se debe considerar que las peticiones del usuario en el mismo, permitan acceder de forma precisa a los datos solicitados, además de lograr que la experiencia de interacción en el mismo sea atractiva. Para obtener los resultados esperados es necesario trabajar en lo siguiente:

- Una estrategia, técnica, modelo o método que permita organizar y/o sistematizar las tareas a los desarrolladores de software y los profesionales de distintas disciplinas que trabajan en conjunto, de manera interdisciplinaria, para el desarrollo de interfaces de software [4].
- Las características y atributos de calidad de artefactos Web tales como usabilidad, navegabilidad, seguridad, características de búsquedas y recuperación de información, performance, accesibilidad, mantenibilidad, no-deficiencia, y aspectos de funcionalidades específicas del dominio que deben ser sistemáticamente tenidas en cuenta durante el proceso de desarrollo de modo de garantizar la calidad en consideración de la satisfacción de los requerimientos de los diferentes perfiles de usuario [5].
- Los criterios del diseño de experiencias de usuario (UX), que son utilizados para los sitios webs, aplicaciones móviles, interfaces de voz, y una serie de dispositivos utilizados en la industria y por los consumidores en su vida diaria. El diseño de UX es un enfoque realista del proceso de diseño de productos digitales que requieren algún tipo de interacción humana. Para los diseñadores, la norma que hay que seguir es intentar que la calidad de la experiencia de los usuarios sea una prioridad máxima [6].
- El empleo de la web semántica para el acceso inteligente y preciso a grandes repositorios de datos [7].
- La implementación de un modelo de datos que incluya información semántica la cual puede ser gestionada mediante sistemas NoSQL [8], [9]. El objetivo de los modelos de datos semánticos es capturar el significado de los datos mediante la integración de conceptos relacionales con conceptos de abstracción más poderosos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se encuentra adecuado a las líneas prioritarias expuestas por la Facultad de Ingeniería de la UNJu en la Resolución FI N° 071/98, en la que se incluye el área temática “Ingeniería de Software”, y se consideran las siguientes líneas de acción: Repositorios digitales, Gestión de la información y el conocimiento, Sistemas de información web, bases de datos y Recuperación de la información. En la actualidad se trabaja en:

- El establecimiento de métricas de calidad web para la evaluación de las principales características de los diferentes repositorios digitales institucionales.
- Definición del algoritmo de búsqueda adecuado al repositorio digital de la Facultad de Ingeniería.
- Aplicación de UX para mejorar la experiencia del usuario en el uso del repositorio institucional.
- Definición del diseño de la base de datos y el gestor de base de datos adecuado al repositorio institucional.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto de investigación tiene como propósito desarrollar una herramienta de búsqueda que facilite el análisis y comprensión de los datos almacenados en el repositorio digital de trabajos finales de grado de la Facultad de Ingeniería de la UNJu.

La herramienta propuesta combinará para su desarrollo, tecnologías de web semántica y sistemas NoSQL. La información extraída de estos repositorios será utilizada como apoyo para la toma de decisiones, tanto a nivel administrativo y operativo de los estudiantes de grado ya que les proporciona el conocimiento necesario para llevar a cabo la selección del tema de trabajo final. Además, esta información permitirá que otros usuarios como egresados, docentes, investigadores y agentes externos conozcan las diferentes líneas de investigación de los trabajos desarrollados, logrando de esta forma la transferencia de la UNJu hacia la comunidad.

El recurso básico del repositorio digital propuesto lo constituyen los trabajos finales de grado, por esta razón es que en primer término se realizó el relevamiento de los datos de estos trabajos finales de grado de la Facultad de Ingeniería de la UNJu. Como resultado de esta actividad se detectó:

- Se cuenta con un ejemplar de los diferentes trabajos finales de grado presentados. No teniendo acceso a mayor información que la publicada en la web de biblioteca de la institución.
- En la página web se encuentra publicada la siguiente información de los ejemplares: título, autor, idioma, editorial, ISBN y biblioteca depositaria, no se dispone por ejemplo, de un resumen del trabajo, lo que hace difícil conocer el contenido del mismo y/o palabras claves asociadas a la temática del trabajo.
- La búsqueda actual de trabajos realizados está limitada a resultados acotados debido a los escasos datos que se almacenan de ellos.

Los resultados exitosos de las búsquedas realizadas en el repositorio propuesto, no sólo dependen del algoritmo de búsqueda elegido, o la navegabilidad en la página web donde se encuentra alojado el repositorio, sino también será fundamental realizar una correcta catalogación de los trabajos finales, definir el esquema de metadatos adecuado y posteriormente organizarlos en colecciones. A partir de esta premisa es que en segundo término, se definieron los metadatos de los trabajos finales, por ejemplo:

- *Datos Personales.* Apellido y Nombre Autor, DNI Autor.
- *Datos Académicos.* Facultad, Universidad, Carrera, Título de grado.
- *Datos del trabajo final.* Título, Año, Lenguaje, Resumen, Palabras claves, Fecha aprobación, Nota aprobación, Director, Codirector, Asesor, Jurado1, Jurado2, Jurado3, Dirección en la web, Propiedad Intelectual.

Cabe destacar que el repositorio deberá contener los datos y metadatos para que la

herramienta de búsqueda logre mejorar los resultados de búsqueda, para esto será necesario solicitar ante las autoridades competentes la incorporación en el reglamento de trabajos finales de grado, que la presentación de estos trabajos se realicen en soporte digital y en formato pdf, respetando un determinado formato e ítems a completar.

Otra de las actividades que se están desarrollando actualmente, es la determinación de la base de datos a emplear con el repositorio de modo de favorecer los resultados de la herramienta de búsqueda. Para ello se está realizando un proceso de comparación de BD considerando y evaluando las siguientes características:

- cantidad estimada de datos a gestionar,
- escalabilidad para el crecimiento futuro del repositorio,
- velocidad y rendimiento,
- seguridad,
- compatibilidad con las tecnologías a utilizar con el repositorio digital,
- interoperabilidad, asegurándose que la base de datos sea compatible con estándares y protocolos comunes,
- capacidad de búsqueda, para garantizar una recuperación eficiente de la información.

En lo que respecta a la producción de trabajos científicos y considerando los objetivos descritos anteriormente, durante el año 2022 se obtuvo como resultado el trabajo "*Estudio preliminar para el desarrollo de una herramienta de búsqueda en repositorios académicos*". Gálvez Díaz, María del Pilar; Martínez, Sergio L.; Cáceres, Nélica R.; Tolaba, Ana C.; Quispe, José R.; Mullicundo, Felipe F.; Villarrubia, Laura R., XXIV Workshop de Investigadores en Ciencias de la Computación (WICC 2022, Mendoza). ISBN: 978-987-48222-3-9. Páginas: 341-344.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está siendo desarrollado por un equipo conformado por docentes investigadores del Grupo de Investigación y

Desarrollo en Ingeniería de Software (GIDIS) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. La estructura del equipo de investigación es la siguiente:

- Director: Mg. Ing. Sergio Luis Martínez. Categoría de Investigación III.

Investigadores:

- Mg. Ing. Nélica Raquel Cáceres. Categoría de Investigación IV.
- Ing. Ana Carolina Tolaba. Categoría de Investigación V.
- Lic. Felipe Fernando Mullicundo. Categoría de Investigación V.
- Mg. Ing. José Rolando Quispe.

Con la realización de este proyecto de investigación se espera la consolidación de los miembros del grupo. Algunos integrantes participaron en la Dirección de Trabajo Final de Carrera denominado “*Coopertino: Aplicación web que asiste a la formación de grupos de estudio y el trabajo en equipo*”. Alumnos: Eduardo Andrés Albornoz y Matías Ramón Ruiz de la Carrera Ingeniería en Informática de la UNJu. Res. FI N° 352/2021. Aprobado en el año 2022 con nota 10 (diez). Este trabajo permitió evaluar la utilización de tecnologías Web.

Mientras que otros integrantes se encuentran en el proceso de asesoramiento de estudiantes en la elaboración de sus trabajos finales de grado y participación en becas referidos a la temática del mencionado proyecto de investigación.

5. BIBLIOGRAFÍA

[1] San Martín, Patricia; Guarnieri, Griselda y Bongiovani, Paola (2014): “Propuesta sociotecnológica para el desarrollo de repositorios de Acceso Abierto adecuados al contexto universitario argentino”. E-Ciencias de la Información, 1-26.

[2] Maenza, R., & Darin, S. (2016). “Universidades abiertas trabajando en la innovación tecnológica y la transparencia”. Revista Internacional Transparencia e Integridad, RITI nro, 2.

[3] Ramírez, M. R., Soto, M. D. C. S., Moreno, H. B. R., Rojas, E. M., Millán, N. D. C. O., & Cisneros, R. F. R. (2019). “Metodología SCRUM y desarrollo de Repositorio Digital”. Revista Ibérica de Sistemas e Tecnologías de Informação, (E17), 1062-1072.

[4] Hotimsky, T., Molina, W., Salgado, C., Peralta, M., & Sánchez, A. (2020). “Diseño UX: Una Guía para Mejorar la Experiencia de los Usuarios haciendo un uso Eficiente de los Recursos Disponibles” en XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz).

[5] Olsina Santos, A. (1999). “Metodología cuantitativa para la evaluación y comparación de la calidad de sitios web”. (Doctoral dissertation, Universidad Nacional de La Plata-Argentina).

[6] Allanwood, G. & Beare P. (2022). “Diseño de Experiencias de Usuario”. Parramón Paidotribo.

[7] Sulé, A., Centelles, M., Franganillo, J., & Gascón, J. (2016). “Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España”. Revista española de documentación científica, 39(1), e121.

[8] Venkatraman, S., Fahd, K., Kaspi, S., & Venkatraman, R. (2016). “SQL Versus NoSQL movement with big data analytics”. International Journal of Information Technology and Computer Science, 8(12), 59-66.

[9] “NOSQL Databases,” Disponible en: <http://nosql-database.org> Acceso: Febrero, 2024.