

Un modelo de recomendación de protocolos experimentales basado en el contexto de uso del usuario

A recommendation model of experimental protocols based on the user's context of use

J. F. Muñoz-Fernández  ; J. A. Guzmán-Luna 

DOI: <https://doi.org/10.22517/23447214.24817>

Artículo de investigación científica y tecnológica

Abstract— In this article we propose a model of recommendation of experimental protocols based on the user's context of use. As a case study, 107 experiments related to 5 experimental protocols of the soil mechanics laboratory tests of the Civil Engineering in the ASTM (American Society for Testing and Materials) standards are analyzed. Starting from the formal specification of the experimental protocol, a representation model is proposed. This model allows to describe the entities that converge in this scientific product, contributing to the formalization of the experimental protocol in a domain of knowledge that has been little explored from this perspective. With the previous result, a recommendation model is presented. This model takes advantage of the concept of the context of use to highlight within the experimental protocol, the entities that characterize the context in which a user performs a new experiment, reproduces, repeats or audits previously performed experiments. This also contributes to considering the recommendation of this type of products that is not yet visible in the repositories of experimental protocols explored in this research. With two previous models, a software prototype is built and using test cases, the representation model and the recommendation model are validated with satisfactory results in a content-based strategy to achieve a recommendation to the user.

Index terms— ASTM, experimental protocol, recommendation systems, soil mechanics laboratory tests.

Resumen—En este artículo se presenta un modelo de recomendación de protocolos experimentales basado en el contexto de uso del usuario. Como caso de estudio se analizan 107 experimentos relacionados con 5 protocolos experimentales de los ensayos de laboratorio de mecánica de suelos de Ingeniería Civil que están normalizados en los estándares ASTM (American Society for Testing and Materials). Partiendo de la especificación formal del protocolo experimental, se propone un modelo de representación que permite describir las entidades que convergen en este producto científico, contribuyendo así, a la formalización de los protocolos experimentales en un dominio de conocimiento que ha sido poco explorado desde esta perspectiva. Con el resultado anterior, se presenta un modelo de recomendación que aprovecha el concepto del contexto de uso para destacar dentro del

protocolo experimental, las entidades que caracterizan el contexto en el que un usuario realiza un nuevo experimento, reproduce, repite o audita experimentos previamente realizados. Con esto también se contribuye a considerar la recomendación de este tipo de productos, propuesta que aún no aparece visible en los repositorios de protocolo experimentales explorados en esta investigación. Con los resultados anteriores se construye un prototipo de software que, mediante casos de prueba, valida el modelo de representación y el modelo de recomendación con resultados satisfactorios en una estrategia basada en contenido para lograr una recomendación al usuario.

Palabras clave— ASTM, laboratorios mecánica de suelos, protocolo experimental, sistemas de recomendación,

INTRODUCCIÓN

CIENCIA 2.0 es un modelo emergente de conceptos y principios relacionados con la diseminación del conocimiento científico. Desde este concepto, se da el movimiento de los mecanismos tradicionales de la diseminación del conocimiento científico basada inicialmente en manuscritos en papel o en su comúnmente adoptada forma digital, a mecanismos que aprovechen el uso de las nuevas tecnologías, preferiblemente aquellas que están basadas en la infraestructura de Internet. Este nuevo enfoque permite compartir y publicar de manera abierta no solo el producto final de la investigación, sino también de todos los demás productos y elementos que soportan y evidencian la evolución del proceso científico y que “son relevantes para su interpretación, evaluación y reutilización de total de la actividad o parte de ella” [1].

En esta nueva dinámica de la diseminación de los productos científicos, el protocolo experimental (P.E) es un elemento fundamental que soporta y facilita la comprensión de los resultados obtenidos y permite la reproducibilidad de un experimento detallando las entidades convergentes en los

Este manuscrito fue enviado el 10 de junio de 2021 y aceptado el 21 de junio de 2022

J. F. Muñoz-Fernández, Esp. MSc. Docente ocasional, Departamento de Ciencias de la Computación y la Decisión. Universidad Nacional de Colombia, sede Medellín. jfmunozf@unal.edu.co.

J. A. Guzmán-Luna, PhD. Profesor titular, Departamento de Ciencias de la Computación y la Decisión. Universidad Nacional de Colombia, sede Medellín. jaguzman@unal.edu.co.



procesos que componen un experimento científico [2], [3].

Destacando la importancia de este producto científico, se ha dado la evolución hacia nuevos mecanismos de representación computacional que son aprovechables en el contexto de Ciencia 2.0. Iniciativas como SMART Protocols [4] Protocols.io¹ [5], MyExperiment² [6] o ROHub³ [7] ponen de manifiesto esta intención, permitiendo además que todos los elementos de información requeridos para repetir el experimento estén presentes convirtiéndose esto, en un criterio para la reusabilidad y la adaptación del protocolo experimental [3].

Uno de los desafíos en la infraestructura computacional que enmarca Ciencia 2.0 tiene que ver con los aspectos concernientes a cómo el conocimiento científico de los productos (p. ej.: el P.E) de la actividad experimental, pueden ser almacenados, recuperados, reutilizados y recomendados. Este último aspecto (la recomendación) es una estrategia que permite reducir el espacio de información en estas infraestructuras computacionales, permitiendo al usuario filtrar este espacio con base en criterios personales que extienden la recuperación de información más allá de una expresión de búsqueda.

Desde este punto de vista, en este artículo se presenta un modelo de recomendación de protocolos experimentales (P.Es), tomando como caso de estudio, 107 experimentos enmarcados en el quehacer experimental que indican los estándares ASTM en cinco P.Es de mecánica de suelos del dominio de la ingeniería civil.

Constituye un interés particular para nuestra propuesta, el P.E como artefacto de conocimiento ya que en la revisión de trabajos previos y de repositorios de P.Es, no emergen aspectos como: (i) la recomendación como estrategia personalizada de recuperación de información (con excepción de ROHub [7]), (ii) los repositorios de este tipo de artefactos digitales no consideran la actividad experimental del dominio de la ingeniería civil, por lo que se reconoce una oportunidad inicial para trabajar en la formalización de P.Es de este dominio y (iii) en los casos de estudio analizados, el registro de la actividad experimental se hace a mano o en el mejor caso, en documentos de Word y Excel dificultando su representación formal unificada, como artefacto digital recuperable por un modelo de recomendación personalizada al usuario.

Se propone partir de la representación formal unificada de estos P.Es, para permitir trascender el registro de la actividad experimental de las notas tomadas a mano o almacenadas en documentos digitales, a un modelo donde sean claramente identificables el documento del P.E, los experimentos que derivan de éste y las entidades que convergen en el desarrollo de un experimento.

Basado en el modelo de representación anterior, se propone un modelo de recomendación personalizada al usuario explorando el concepto del contexto de uso (C.U). Este C.U permite dar valor no solo al P.E completo sino a los elementos individuales que convergen en él en su propósito de uso que es permitir a un humano realizar un nuevo experimento, reproducir, repetir o auditar experimentos previamente realizados. Con base en los dos modelos anteriores, se implementa un prototipo de software que permite recuperar la

información de los protocolos experimentales y sirve a los propósitos de implementación y validación del modelo de recomendación propuesto.

Basado en lo anterior. Este documento está organizado de la siguiente manera: en la sección dos se describe la fundamentación teórica que soporta los modelos a proponer. En la sección tres se describe la metodología. En la sección cuatro se señalan los trabajos previos con respecto a la problemática planteada. En la sección cinco se presenta el modelo formal de representación. En la sección seis se presenta el modelo de recomendación. En la sección siete se presenta la validación del modelo. Finalmente, en la sección ocho se presentan las conclusiones y propuestas de trabajo futuro.

I. FUNDAMENTACIÓN TEÓRICA

En esta sección se describen los siguientes conceptos que soportan la solución propuesta: el protocolo experimental, los sistemas de recomendación y el contexto de uso.

A. Protocolo experimental (P.E)

El protocolo experimental (P.E) es un instrumento que formaliza el quehacer experimental, describiendo las particularidades del diseño de un experimento. En este diseño del experimento se evidencia la concurrencia de entidades como los actores humanos, sujetos de investigación, materiales de laboratorio, instrumentación, material bibliográfico, recursos de tecnología, entre otras entidades que caracterizan la interdependencia cooperativa del conocimiento científico a través de la actividad experimental [2].

Este instrumento es el corazón de las buenas prácticas de laboratorio y de manufactura, incluso, es esencial para la generación de datos que son requeridos por agencias reguladoras y es el instrumento que permite la reproducibilidad y la confiabilidad de los experimentos como piedra angular de la práctica científica [8].

En ingeniería civil la actividad experimental requerida en proyectos de construcción está descrita en normas o estándares como lo es el estándar ASTM⁴, el cual, aunque no es el único, consiste en un conjunto de documentos orientados a estandarizar las diferentes pruebas de laboratorio describiendo las actividades experimentales, el orden de cada una, las muestras, especímenes, modelos matemáticos, equipos, entre otras entidades, que se requieren para medir y describir las propiedades de los materiales usados en la construcción.

B. Sistemas de recomendación (S.Rs)

Con la adopción generalizada de Internet como una red que ofrece un número significativo de múltiples recursos de información en diferentes dominios de interés y conocimiento, los desafíos para encontrar y reutilizar aquellos recursos precisos ajustados a los requerimientos y preferencias de un usuario han captado el interés de la investigación por desarrollar modelos que faciliten esta labor. Algunos de estos modelos son evidentes en los sistemas de recomendación cuyo propósito principal es ayudar al usuario en la toma de decisiones cuando éste enfrenta un sin número de opciones de posibles de recursos

¹ <https://www.protocols.io/>

² <https://www.myexperiment.org>

³ <http://www.rohub.org/>

⁴ <https://www.astm.org/>

de información, sugiriéndole aquellos recursos que son más ajustados a sus preferencias personales.

Uno de los sistemas de recomendación es el basado en el filtrado de contenido, el cual, requiere conocer con anticipación el perfil del usuario y las características de los ítems que serán recomendados [9]. El perfil del usuario se puede construir explícitamente solicitando al usuario información acerca de sus gustos, intereses o preferencias [10], o también puede ser inferido por el sistema a través del comportamiento que el usuario ha seguido en el pasado haciendo uso del sistema, p. ej.: calificando positivamente algunos ítems o agregándolos a su lista de favoritos. En este modelo, el proceso de recomendación consiste en emparejar los atributos del perfil del usuario con los atributos que describen los ítems en el sistema [11]. Así, un gran espacio de ítems disponibles es reducido a un espacio considerable de opciones que se ajustan al perfil del usuario. Esto facilita la difícil tarea de elegir el ítem adecuado que representa el interés del usuario.

C. Contexto de uso (C.U)

El C.U está asociado al concepto de usabilidad de un producto y por lo tanto los productos deben ser diseñados para contextos específicos de uso. El concepto de usabilidad se refiere al criterio, conjunto de criterios (eficacia, eficiencia y satisfacción) y sus correspondientes métricas (cuantitativas o cualitativas) empleadas para evaluar la calidad de un diseño de un producto [12]–[14]. De acuerdo con la norma ISO 9241-11 [12], el C.U son los “usuarios, tareas, equipamiento (equipo, programas y documentos) y entorno físico y social en el que un producto es utilizado”.

En general el C.U de un objeto está definido por la caracterización de los usuarios, sus objetivos, las tareas que se llevan a cabo para lograr el objetivo, las herramientas que apoyan los objetivos del usuario y el entorno técnico, social y organizacional en el cual se da el uso del objeto. El objeto bien podría ser un producto específico como se señala en [12]–[14], o un sistema como en el caso indicado por [15]–[17]. En cualquier situación, los elementos del contexto de uso donde converge su caracterización son: el usuario, el objeto que se propone utilizar y los objetivos con los que el usuario usa el objeto.

II. METODOLOGÍA

En el contexto metodológico de esta investigación se emplearon dos puntos de vista: (i) el teórico y (ii) el aplicado. Desde el primero se identifica el concepto del P.E, su importancia para la diseminación científica asociada a la reproducibilidad de los experimentos y la formalización como elemento de conocimiento explotable computacionalmente. Así mismo, se logra establecer un marco de referencia para los P.Es de ingeniería civil, particularmente aquellos que están estandarizados bajo las normas ASTM asociados a los laboratorios de mecánica de suelos, en donde se identifican sus elementos y se establece un criterio para lograr su representación aprovechable en un modelo de recomendación basado en el C.U del usuario. En esta misma dirección se realizó una revisión de literatura concerniente a modelos de recomendación para P.Es identificando no solo los pormenores de las propuestas de otros investigadores, sino también las

oportunidades de contribución a través del análisis comparativo. Finalmente, se exploró el concepto de C.U y los modelos de recomendación que lo aprovechan para resolver los problemas de sobrecarga de información en diferentes dominios de conocimiento para identificar cómo podría aprovecharse este concepto en la recomendación de protocolos experimentales (P.Es).

Desde la perspectiva aplicada, se realiza la implementación del modelo de recomendación, en un prototipo de software que permitió su validación y la evaluación de sus fortalezas y debilidades.

III. TRABAJOS PREVIOS

En esta sección se resume la revisión de trabajos previos con relación a cuatro aspectos: la representación de P.Es; los S.Rs y el C.U; la exploración de los repositorios de P.Es con relación al dominio de conocimiento (Ing. Civil) y a los modelos de recuperación o recomendación.

A. Representación de protocolos experimentales (P.Es)

La Web se ha vuelto una plataforma de divulgación científica que ha ampliado la diversidad de recursos digitales disponibles para la investigación [6]. Aprovechando las tecnologías de Web 3.0 en una infraestructura más inteligente [18] en donde se pretende que la información tenga un significado tanto para las máquinas como para los humanos, se han dado diferentes esfuerzos en la representación de los P.Es a través de las ontologías. Una ontología se define como “una especificación explícita y formal de una conceptualización compartida” [19].

En la Tabla I se resume el dominio y las características principales de seis ontologías que se usan en la representación de P.Es. En esta revisión se puede observar que el dominio de ingeniería civil se ha abordado en una sola propuesta (Sciex [25]). Sin embargo, esta propuesta solo considera en la representación de la secuencia de las actividades experimentales, los aspectos computacionales de cálculos matemáticos que describen los documentos ASTM y no se representa la retórica de otros elementos del procedimiento de laboratorio del P.E.

TABLA I
REVISIÓN DE ONTOLOGÍAS PARA LA REPRESENTACIÓN DE P.ES

Ontología	Dominio						Características						
	GE	BM	BO	BT	ME	ES	IC	A	S	L	M	P	T
[20]	X									X	X		X
[8]		X											X
[21]			X		X				X				
[22]			X		X							X	
[23]				X				X	X			X	X
[24]						X			X				X
[25]							X		X		X	X	X

GE: General. **BM:** Biomedicina. **BO:** Biología. **BT:** Biotecnología. **ME:** Medicina. **ES:** Experimentación *in silico*. **IC:** Ingeniería civil. **A:** Actualización permanente. **S:** Representa la secuencia de actividades. **L:** Permite la expresividad en lenguaje natural. **M:** Representa formalismos matemáticos. **P:** Representa los elementos del P.E. **T:** Representa actividades del P.E.

Algunas otras limitantes, con excepción de [23], incluyen la falta de actualización de la ontología

B. Sistemas de recomendación (S.Rs) y contexto de uso (C.U)

En la Tabla II se resume la revisión de los sistemas de recomendación sensibles al contexto con base en las dimensiones de análisis del contexto de uso que convergen en las definiciones de [12]–[15]. Estas dimensiones incluyen la caracterización del usuario (tipos de usuario, conocimientos, habilidades, experiencia, características personales), dimensión de las tareas, entorno organizativo, entorno social, entorno ambiental y tiempo.

TABLA II
S.RS QUE CARACTERIZAN LAS DIMENSIONES DEL C.U

Dimensiones del contexto de uso/Trabajos revisados	TU	CHE	CP	TA	EOS	ETD	EFA	T
[26]			X	X	X			
[27]				X				
[28]			X	X			X	X
[16]			X	X	X	X		
[29]			X	X	X	X	X	
[17]		X	X			X	X	
[30]		X	X	X	X	X	X	
[12]	X	X		X		X		
[13]	X	X		X		X		
[31]			X	X			X	X
[32]		X	X	X				
[14]	X	X		X		X		
[33]			X	X				
[34]			X	X				
[15]	X		X		X	X	X	
[35]				X				
[36]			X					
[37]				X				
[38]			X		X	X		

TU: Tipos de usuarios. **CHE:** Conocimientos habilidades y experiencia. **CP:** Características personales. **TA:** Tareas/Actividades. **EOS:** Entorno organizativo/social. **ETD:** Entorno técnico, equipamiento o dispositivo. **EFA:** Entorno físico o ambiental. **T:** Tiempo.

En la Tabla II se destaca que la dimensión de las tareas es un elemento que se tiene en cuenta en S.Rs sensibles al contexto.

C. Repositorios de protocolos experimentales (P.Es)

Debido a la falta de información concreta en relación con sistemas o modelos de recomendación de P.Es, las Tablas III y IV resumen las características de un trabajo exploratorio que consistió en: (a) registrarse en diferentes plataformas de P.Es (y productos subyacentes) y (b) visitar la plataforma e interactuar con los elementos disponibles tanto a nivel del P.E (visitar/explorar) como a nivel de la interfaz del usuario (explorar/buscar) con la intención de identificar si emergía el dominio de ingeniería civil en los objetos almacenados en el

repositorio explorado y de identificar modelos/estrategias de recuperación y/o recomendación.

En la Tabla III se puede evidenciar que los repositorios de P.Es explorados, ninguno almacena P.Es del dominio de ingeniería civil. En la Tabla IV se evidencia que solo uno de los repositorios implementa un S.R basado en el contenido. Con la particularidad que dicho repositorio es de Objetos de Investigación [58] y no de P.Es. Los demás repositorios implementan estrategias de búsqueda y recuperación basada en áreas temáticas, palabras clave, autores, resumen, entre otros metadatos.

TABLA III
REPOSITORIOS P.ES CON RELACIÓN AL DOMINIO DE CONOCIMIENTO

R/D	CN	FT	ME	QU	MA	CO	IB	BM	MU	BI	IN	BF	BE	BF	SI	FI
[5]	X	X	X	X												
[39]	X		X		X	X										
[40]							X			X						
[41]	X															
[42]											X					
[43]								X								
[44]									X							
[45]	X															
[46]	X							X								
[47]								X		X						
[48]											X					
[49]												X	X			
[50]										X				X		
[51]																X
[52]	X															
[53]																
[54]	X									X						
[55]	X		X													
[56]									X							
[57]			X	X						X						X
[6]														X		
[7]									X							

R/D: Repositorio/Dominio. **CN:** Ciencias Naturales. **FT:** Flujo de trabajo computacional. **ME:** Medicina. **QU:** Química. **MA:** Matemáticas. **CO:** Computación. **IB:** Ingeniería Biológica. **BM:** Biología Molecular. **MU:** Multidominio. **BI:** Biología. **IN:** Investigación Neuromuscular. **BF:** Biofísica. **BE:** Biomedicina. **BF:** Bioinformática. **SI:** Sin información. **FI:** Física.

La exploración de los repositorios que describe la Tabla IV demuestra que la comunidad científica se ha esforzado en disponer de diferentes infraestructuras para trabajar en comunidad, compartir el conocimiento y descubrir los productos de la actividad de investigación.

Se destaca que el dominio de ingeniería civil es un dominio que aún no aparece en estos repositorios y que en dichos

repositorios no emerge un modelo de recomendación personalizada al usuario, con excepción del caso mencionado.

TABLA IV
REPOSITORIOS P.ES CON RELACIÓN A LA ESTRATEGIA DE RECUPERACIÓN/RECOMENDACIÓN

	Búsqueda y recuperación											R					
	Áreas temáticas	Palabras clave	Lugar	Fechas	Autor	Resumen	Título	Técnica del protocolo	Organismo	Personas/Usuarios/Grupos	Proyectos		Investigaciones	Publicación	Institución	Archivos de datos	Sugerencias
[5]	X																X
[39]		X		X	X	X	X										
[40]	X																X
[41]	X		X				X	X									X
[42]									X	X	X				X	X	
[43]	X																
[44]	X																
[45]																	
[46]	X																
[47]	X																
[48]	X																
[49]	X																
[50]	X																
[51]	X																
[52]	X																
[53]	X																X
[54]																	X
[55]		X	X									X	X				
[56]	X	X															X
[57]	X			X	X									X			
[6]	X	X							X								X
[7]	X																X

R: Modelo de recomendación.

IV. MODELO DE REPRESENTACIÓN

Para precisar la dinámica y el contexto de los P.Es experimentales en los laboratorios de ingeniería civil, hay que comprender la relación entre el documento del P.E, la persona que lleva a cabo el experimento y el resultado de esta relación. El documento del P.E indica a las personas en el laboratorio cómo hacer el experimento; las personas orquestan los recursos necesarios del laboratorio, su saber y experiencia para llevar a

buen término el experimento. Durante la elaboración del experimento, se lleva un registro de toda la actividad experimental a modo de cuaderno de laboratorio y una vez el experimento ha finalizado se produce un reporte con los resultados obtenidos del ensayo.

En Fig. 1 (documento del P.E) se ilustra un extracto de la sección del procedimiento de laboratorio y cálculos matemáticos que se indican en el P.E (documento del P.E) para determinar el contenido de humedad gravimétrica de una muestra de suelo con base en el estándar ASTM D2216 [59] y tomado de [60]. En el discurso de esta sección del documento del P.E se identifican las acciones, y los recursos necesarios para llevar a término cada actividad individual del procedimiento de laboratorio. Este procedimiento describe un flujo de trabajo de actividades ordenadas secuencialmente: flujo de trabajo experimental.

En Fig. 2 (cuaderno de laboratorio) se ilustra el registro de las actividades experimentales ejecutadas por la persona que lleva a cabo el experimento y que corresponden a las actividades indicadas en el documento del P.E de la Fig.1

El análisis del documento del P.E y del cuaderno de laboratorio son el punto de partida para proponer un modelo de representación que permita: (a) representar el documento del P.E con todas las secciones del documento y (b) representar el cuaderno de laboratorio (experimento) que corresponde a un experimento puntual, detallando por cada actividad individual del flujo de trabajo experimental: entradas, salidas y recursos usados

Procedimiento

- Seleccione la muestra y los tres especímenes a los que se les calculará el contenido de humedad.
- Pese y registre el peso de las taras: (t)
- Coloque cada espécimen en una tara y registre el peso de la tara más la muestra de suelo húmeda (Wh + t)
- Lleve la tara con la muestra húmeda al horno y seque a una temperatura entre 105°C y 110 °C, hasta peso constante. En la mayoría de los suelos se requiere entre 12 y 16 horas para obtener esta condición. Para mayor seguridad, tome pesos sucesivos de la tara, más la muestra de suelo seco, con diferencia de 1 hora, hasta que no haya cambio de peso entre cada pesaje.
- Retire del horno la tara y la muestra de suelo seca y deje enfriar hasta que pueda manipular con las manos.
- Pese y registre el peso de la tara más la muestra de suelo seco (Ws+t).

Cálculos

- Determine el peso del agua en los poros (Ww), como la diferencia entre el peso del suelo húmedo menos el peso del suelo seco:

$$Ww = (Wh + t) - (Ws + t) \tag{1}$$

- Determine el peso de los sólidos (Ws) como el peso del suelo seco:

$$Ws = (Ws + t) - t \tag{2}$$

Fig. 1. Extracto de la sección de procedimiento en el documento del P.E: Actividades individuales del flujo de trabajo experimental organizadas secuencialmente. Tomado de [60].

Fecha: 12-04-2016
 Cliente: Proyecto A
 Muestra: SP-23 – Proyecto A.
 Autor: Juan Pérez

- Registro del experimento para determinar el contenido de agua en una muestra de suelo.
1. Se seleccionaron los tres especímenes de acuerdo con la muestra de suelo suministrada. Se usó la muestra de suelo correspondiente al proyecto A. Se etiquetan los especímenes como ESP-A1, ESP-A2 y ESP-A3 respectivamente.
 2. Se pesan en la balanza del laboratorio las taras vacías. Se registran los pesos de las taras de los especímenes en el archivo de resultados del experimento.
 3. Se pesan en la balanza del laboratorio las taras con los especímenes húmedos. Se registran los pesos de las taras de los especímenes en el archivo de resultados del experimento para cada espécimen del ensayo.
 4. Se llevan las taras con las muestras húmedas al horno del laboratorio. Se aplican 105°C para el secado de las muestras. Espécimen de suelo se dejan secando 16 horas. Proceso inicia a las 15:30 horas.
 5. Se pesa y se registra el peso de cada tara más los especímenes de suelo seco. Se anotan estos pesos en el archivo de resultados.
 6. Se calcula el peso del agua con base en la ecuación $W_w = (W_h + t) - (W_s + t)$ se deja registro de los cálculos en el archivo de resultados del experimento.
 7. Se calcula el peso de los sólidos con la ecuación $W_s = (W_s + t) - t$ el resultado del cálculo se deja anotado en el archivo de resultados del experimento.
 8. Se calcula el contenido de humedad gravimétrica con la ecuación: $W(\%) = \frac{W_w}{W_s} \times 100$ Se deja registro del resultado en el archivo de resultados del experimento.

Fig. 2. Cuaderno de laboratorio (experimento) para el P.E de la Fig 1. Registro de lo que hace la persona que lleva a cabo el experimento.

En Fig. 3. Se ilustra una visión de la riqueza y complejidad de elementos que convergen en un cuaderno de laboratorio (experimento) y que podrían ser aprovechables en un modelo de recomendación personalizada al usuario. Por ejemplo, en el centro de la figura se encuentra el flujo de trabajo experimental que describe el documento del P.E en el procedimiento de laboratorio. Este flujo de trabajo indica paso a paso como llevar a cabo el experimento, y es en este flujo de trabajo, donde convergen todas las demás entidades que relaciona el documento del P.E. Estas otras entidades son: materiales y equipos, archivos de reportes de resultados, software, autores, instituciones y laboratorios.

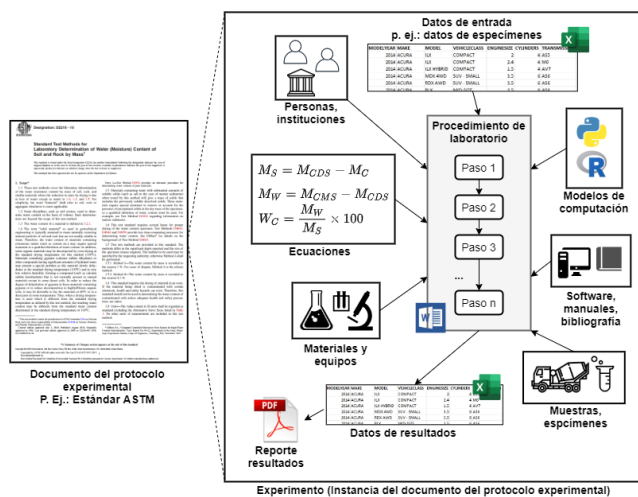


Fig. 3. Elementos que convergen en la realización de un experimento registrado en un cuaderno de laboratorio. El documento del P.E da origen a un experimento puntual.

A. Representación del documento del protocolo experimental

Una de las dificultades observadas en los documentos de los P.Es, es que su contenido varía en los siguientes aspectos: nombrado de las secciones de documento a documento, número de secciones, secciones presentes/ausentes en cada documento. Por esta razón, en la representación del documento P.E se propone un modelo que permita a los autores registrar el contenido del documento en el número y nombrado de las secciones que se necesiten. En Fig. 4 se ilustra un diagrama de clases UML como modelo de representación del documento del P.E.

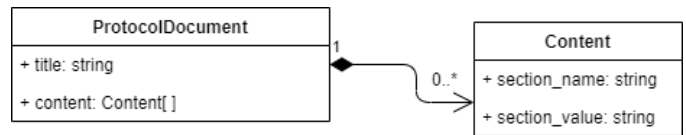


Fig. 4. Modelo que representa a un documento de un protocolo experimental como una agregación de dos clases.

El diagrama de clases de la figura anterior indica que un documento está formado por dos objetos: ProtocolDocument y Content. El primero tiene como atributos al título del documento (title) y al contenido (Content[]). Este último representado como un arreglo de objetos de tipo Content. Así, el objeto Content representa el nombre de una sección del documento (section_name), p. ej., “Objetivos” y el contenido textual (string) de dicha sección (section_value).

B. Representación del experimento (cuaderno de laboratorio)

Para la representación del experimento el modelo de representación considera los siguientes aspectos: (a) las actividades individuales del flujo de trabajo experimental, (b) los recursos que participan y el rol (entrada o salida) de cada actividad experimental (p. ej.: materiales y equipos, suministros, ecuaciones, muestras, especímenes, archivos, software), (c) las terceras partes como instituciones, laboratorios, y universidades que participan en el experimento, (d) los autores y (e) el vínculo que existe entre el documento del P.E y el experimento.

En Fig. 5 se ilustran las clases con las que se representa el experimento y su vínculo con el documento del P.E representado en las clases de Fig. 4.

La clase Experiment permite representar al experimento con información básica. La clase Procedure permite representar las actividades individuales del flujo de trabajo experimental. La clase Resource permite representar los diferentes recursos que participan como entradas y salidas en cada actividad individual del flujo de trabajo experimental. La clase InputOutput permite representar el rol de un recurso. La clase ThirdPartie permite representar las terceras partes que participan en el experimento. La clase Author permite representar a los autores del experimento. Las subclases de *Attribute permiten describir a los recursos, autores y terceras partes con todos los atributos y valores necesarios.

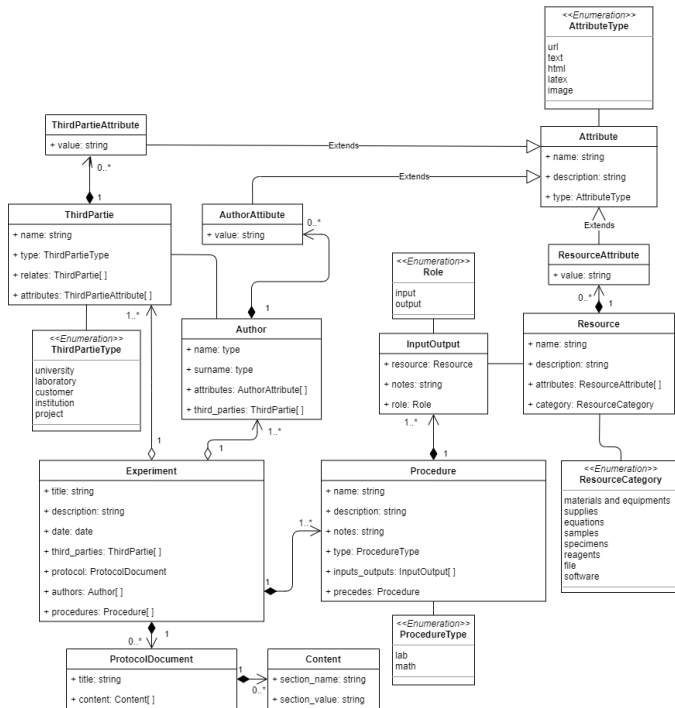


Fig. 5. Modelo completo de la representación del experimento y su relación con el documento del protocolo experimental.

El modelo completo de Fig. 5 parte de los análisis que propone SMART Protocolos [4]. Sin embargo, en [4] se propone una ontología ajustada a los P.Es en ciencias naturales. Nuestro modelo se propone para la representación de P.Es en un dominio diferente al de las ciencias naturales, el cual corresponde al de ingeniería civil. Dominio de conocimiento que hasta ahora solo es visible en la propuesta de la ontología Scienex [25].

También se propone el uso de información de tipo *string* de para algunos de los atributos del modelo con la intención de obtener una representación cercana al lenguaje natural, con la expresividad suficiente con la que un humano puede crear nuevos P.Es y documentarse para repetir, reproducir, auditar experimentos, así como para registrar toda la actividad experimental de un ensayo de laboratorio.

V. MODELO DE RECOMENDACIÓN

Teniendo presente la importancia que tienen los elementos individuales en la usabilidad de los P.Es y de los experimentos derivados de ellos, se propone un modelo de recomendación que, aprovechando el modelo de representación descrito en la sección IV, sean considerados sus elementos en el proceso de recomendación personalizada al usuario.

Así mismo, se propone caracterizar el C.U para un modelo de recomendación de P.E en dos dimensiones: (a) tareas y (b) entorno técnico. En la Tabla V se detallan los elementos que hacen parte de la caracterización de cada una de estas dimensiones del C.U.

TABLA V
CARACTERIZACIÓN DEL CONTEXTO DE USO (C.U) DEL P.E

Tareas	Entorno Técnico		
Actividades individuales del flujo de trabajo experimental	Materiales y equipos	Ecuaciones	Documento del P.E
Descripción de los atributos usados en la caracterización			
Nombre, descripción, título experimento, nombre de todos los recursos que participan como entradas y salidas de la actividad	Nombre del recurso y descripción del recurso	Nombre del recurso, descripción del recurso y expresión en LaTeX de la ecuación	Título del documento, contenido de cada sección
Atributos del modelo de representación Fig. 5			
P.name	R.name	R.name	PD.title
P.description	R.description	R.description	C.section_value
E.title		RA.value	
R.Name			

P: clase Procedure. **E:** clase Experiment. **R:** clase Resource. **RA:** clase ResourceAttribute. **PD:** clase ProtocolDocument. **C:** clase Content (ver Fig. 5)

Ya que el usuario también hacer parte del análisis dimensional del C.U, este se caracteriza a partir del perfil del usuario (P.U) construido con base en las preferencias explícitas que este hace empleando un mecanismo de calificación unaria (agregar a una colección de favoritos) sobre los elementos de la tabla anterior. Estas preferencias del usuario sobre los elementos del experimento constituyen su P.U y éste a su vez describe el C.U en el que el usuario realiza un nuevo experimento, reproduce, repite o audita un experimento realizado previamente.

Aprovechando la naturaleza textual del modelo de representación y teniendo en cuenta la caracterización del C.U, para el modelo de recomendación se usa un algoritmo basado en contenido. Para ello se propone representar los elementos o ítems del C.U descritos por la Tabla V, mediante un modelo de espacio vectorial de palabras clave, haciendo uso del cálculo de una matriz de ponderado TF-IDF [9], [11] en donde cada ítem o elemento se representa como un vector fila (de la matriz) de n-dimensiones y cada dimensión corresponde a un término o palabra-clave ponderada en el espacio n-dimensional que está compuesto por todo el conjunto de ítems o elementos [11].

Teniendo en cuenta lo anterior, se obtienen cuatro matrices TF-IDF_{n×m} que representan al C.U indicado por la Tabla V.

- TF-IDF_{procedimientos}: representa todas las actividades individuales del flujo de trabajo experimental de todos los experimentos.
- TF-IDF_{recursos_me}: representa todos recursos de la categoría de los materiales y equipos de todos los experimentos.
- TF-IDF_{ecuaciones}: representa todos los recursos de tipo ecuación usadas en los experimentos.
- TF-IDF_{documentos_pe}: representa a todos los documentos de los P.E.

Ya que el usuario puede especificar una preferencia explícita por los elementos del C.U, se requieren cuatro vectores pivote, uno por cada matriz TF-IDF, que representa la preferencia

(indicada como 1) o la no preferencia (indicada como 0) del usuario por los elementos del C.U. Los cuatro vectores de Preferencias_{1×m} son los siguientes: (a) Preferencias_{procedimientos}, (b) Preferencias_{recursos_me}, (c) Preferencias_{ecuaciones} y (d) Preferencias_{documentos_pe}.

Con las matrices TF-IDF y los vectores de preferencias, se construye el perfil de usuario (P.U) a través del producto escalar entre el vector de preferencias y la matriz TF-IDF correspondiente, obteniendo cuatro vectores de P.U, uno por cada C.U:

- $TF-IDF_{procedimientos} \times Preferencias_{procedimientos} = PU_{procedimientos}$
- $TF-IDF_{recursos_me} \times Preferencias_{recursos_me} = PU_{recursos_me}$
- $TF-IDF_{ecuaciones} \times Preferencias_{ecuaciones} = PU_{ecuaciones}$
- $TF-IDF_{documentos_pe} \times Preferencias_{documentos_pe} = PU_{documentos_pe}$

Una vez se han obtenido los cuatro vectores P.U, se calcula el coeficiente de similitud (C.S) entre el vector P.U, con cada una de las filas de la correspondiente matriz TF-IDF. Este cálculo se realiza mediante la ecuación de similitud coseno (1).

$$sim(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Donde, A corresponde a cada vector fila de la correspondiente matriz TF-IDF y B corresponde al vector perfil del usuario correspondiente con cada matriz TF-IDF. Además $sim(A, B) = [0, 1]$, donde 1 indica que el vector fila representado en la matriz TF-IDF corresponde con el gusto o preferencia del usuario y 0 indica todo lo contrario.

Es importante mencionar que la recomendación que ofrece el modelo al usuario es una lista de experimentos en donde aparecen los elementos que registran mayores C.S ya que estos elementos son los que caracterizan su C.U en la configuración de nuevos experimentos o en la repetición, reproducción y auditoria de experimentos previamente realizados. Para lograr esta recomendación final se tiene en cuenta lo siguiente:

- Se ordenan los vectores de C.S de mayor a menor C.S. Se obtiene un Top-N de elementos en cada caso: Top-N_{procedimientos}, Top-N_{recursos_me}, Top-N_{ecuaciones} y Top-N_{documentos_pe}. De cada Top-N es necesario descartar aquellos elementos que ya el usuario indicó que son de su preferencia.
- Se buscan los experimentos en donde participan cada uno de uno de los elementos del Top-N. Esto puede dar lugar a experimentos que se repiten. Por ejemplo, una ecuación particular y un recurso de la categoría de materiales y equipos, presentes en el Top-N_{ecuaciones} y en el Top-N_{recursos_me} respectivamente, pueden coincidir en el mismo experimento. Incluso también puede darse el caso que dos elementos diferentes del mismo Top-N den lugar al mismo experimento. En cualquiera de estas dos situaciones se

debe seleccionar una sola ocurrencia.

- Las listas de experimentos que produce cada Top-N, se presenta al usuario con un criterio de relevancia. Por ejemplo, si el usuario ha indicado mediante el mecanismo de calificación, un número significativo de ecuaciones preferentes, quiere decir que, para él, estos elementos son más importantes que, por ejemplo, las actividades del flujo de trabajo experimental en donde no ha indicado preferencia alguna o donde el número de elementos preferentes es menor al de las ecuaciones. En este caso, el criterio de relevancia para presentar experimentos recomendados al usuario que produce cada Top-N, será un conteo de los elementos que el usuario previamente ha indicado que son de su preferencia en cada caso.

En general se propone un modelo de recomendación en donde se resaltan las siguientes características.

- Se considera la importancia de elementos individuales al interior del experimento al permitir al usuario indicar una preferencia explícita por éstos.
- Los elementos sobre los cuales el usuario puede indicar una preferencia explícita están descritos por su propio conjunto de atributos.
- La recomendación depende de la preferencia de estos elementos individuales, pero considerando que estos no existen como elementos aislados, sino que existen al interior y en el contexto del experimento.

A. Arquitectura

En Fig. 6 se ilustra la arquitectura propuesta para la implementación del modelo de recomendación que está compuesta por las siguientes capas.

- Capa de perfiles: Genera la representación (en el modelo del espacio vectorial de palabras clave) de los elementos que participan en los experimentos y genera los diferentes vectores de los P.U que relejan las preferencias del usuario por los elementos individuales de las dimensiones de C.U.
- Capa de datos: permite el acceso a los siguientes repositorios: (a) Protocolos experimentales: almacena los documentos de los P.E y los experimentos derivados de ellos. (b) Usuarios: almacena la información que permite identificar unívocamente al usuario dentro del sistema y los vectores perfil de usuario. (c) Ítems: almacena las matrices TF-IDF de los elementos de las dimensiones de C.U. (d) Preferencias de usuario: registra mediante el mecanismo de calificación unaria, los diferentes elementos de las dimensiones de C.U que el usuario puede indicar como preferencia. (e) Recomendaciones: almacena el cálculo de las recomendaciones para cada usuario en función de cada uno de sus P.U.
- Capa de recomendación: se calculan las recomendaciones con base en las matrices TF-IDF, las preferencias de usuario y los vectores de P.U. El componente de contextualización de la recomendación se encarga de producir los diferentes Top-N que son presentados al usuario.
- Capa de formato de contenido: genera las estructuras de datos que permiten a la capa de aplicación presentar el

contenido de los documentos de los P.Es y de los experimentos.

- Capa de aplicación: esta capa presenta al usuario el contenido del repositorio Protocolos Experimentales. El componente Colector de eventos se encarga de suministrar el mecanismo de calificación unaria. A través de este componente el usuario puede (i) agregar a la colección de elementos preferentes del usuario un elemento del experimento, es decir, crear un registro en el repositorio Preferencias del usuario y (ii) eliminar un elemento de su colección de elementos preferentes que están registrados en el repositorio de Preferencias del usuario.

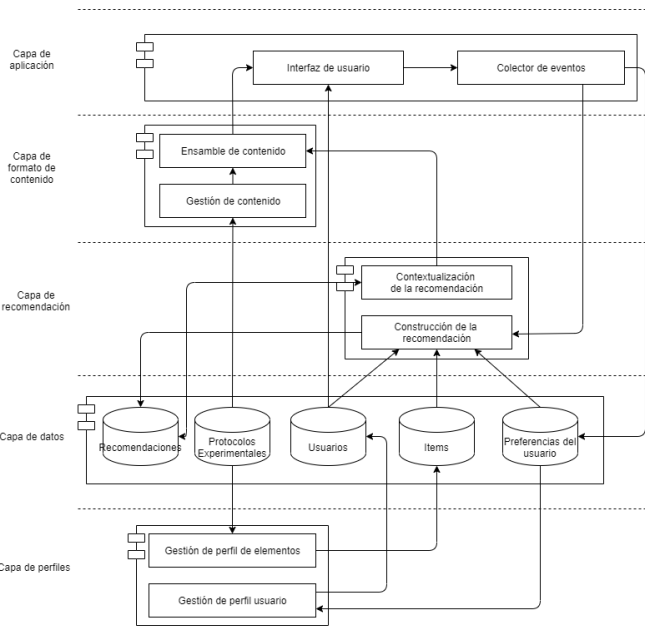


Fig. 6. Arquitectura para la implementación del modelo de recomendación.

B. Implementación

Se construyó un prototipo de software Web (HTML5 + CSS3) usando el lenguaje de programación Python (versión 3.6) y librerías de apoyo como Flask, Numpy, SciPy, Scikit-Learn, Pandas, NLTK. En Fig. 7 se ilustra una captura de pantalla del prototipo de software construido y en Fig. 7 se ilustra el diagrama de implementación.

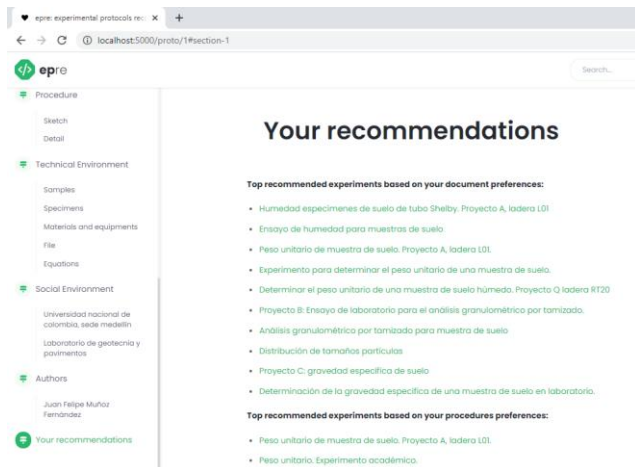


Fig. 7. Captura de pantalla del prototipo de software construido.

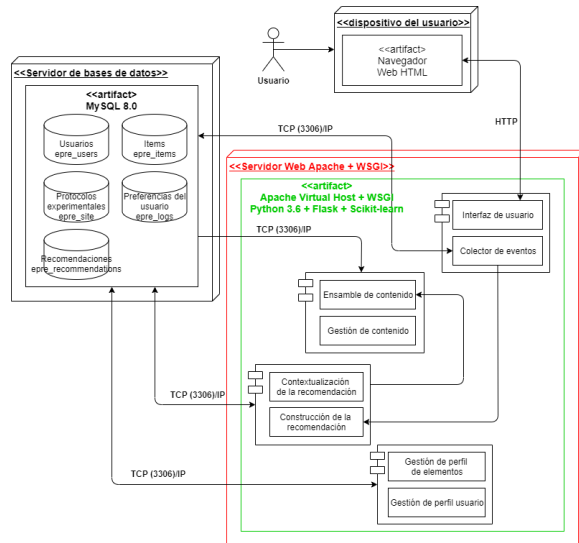


Fig. 8. Diagrama de implementación del prototipo de software desarrollado para la implementación del modelo de representación y de recomendación.

VI. VALIDACIÓN DEL MODELO

Para la validación del modelo de representación y de recomendación se usaron un total de 107 (casos de prueba) experimentos del laboratorio de mecánica de suelos basados en los documentos de los P.E tomados de [60]. Estos 107 experimentos conforman un total de 960 actividades/tareas individuales del flujo de trabajo experimental, 14 ecuaciones, 67 recursos de la categoría de materiales y equipos y seis documentos de los siguientes protocolos experimentales (P.E).

- Guía-LMS-PCJIC-02: determinación en laboratorio del contenido de humedad gravimétrica de suelo.
- Guía-LMS-PCJIC-03: determinación del peso unitario de suelos (con volumen conocido).
- Guía-LMS-PCJIC-04: determinación de la gravedad específica de los suelos.
- Guía-LMS-PCJIC-05: análisis granulométrico por tamizado.
- Guía-LMS-PCJIC-06: análisis granulométrico por hidrómetro.

En Fig. 9 se ilustra la distribución de experimentos por cada uno de los documentos de los protocolos experimentales del listado anterior.

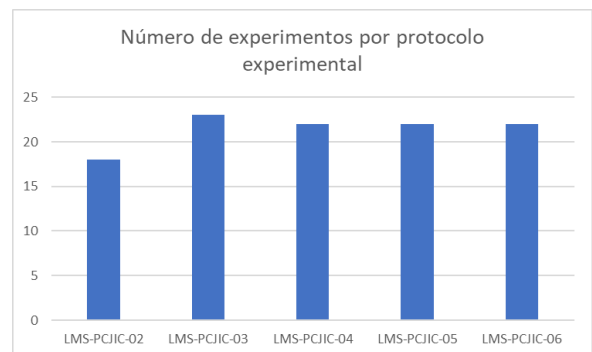


Fig. 9. Número de experimentos (casos de prueba) por documento de protocolo experimental.

Las pruebas que a continuación se describen, se realizaron usando el prototipo de software construido en donde se le pidió a un usuario que seleccionara como preferente, los diferentes elementos del C.U que describe cada prueba para calcular las recomendaciones.

A. Prueba 1: preferencia por un documento del P.E

En esta prueba se miden las preferencias del usuario con relación a los documentos del P.E. almacenados en el repositorio. La Fig. 10 se ilustran los coeficientes de similitud de los documentos de P.Es con el vector P.U: $PU_{documentos_pe}$. Es decir, aquí el usuario ha marcado como preferencia, documentos los documentos de P.Es indicados en la leyenda de la gráfica. Con base en estas preferencias, el modelo calcula la recomendación de experimentos relacionados con estos documentos. Estas recomendaciones son las que se detallan en la Tabla VI.

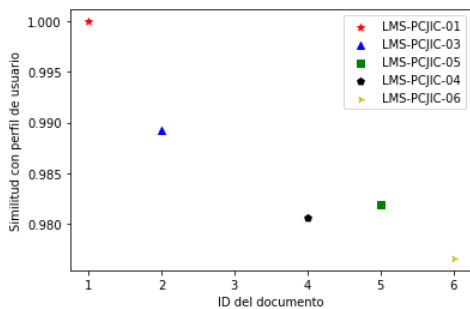


Fig. 10. Coeficientes de similitud de los documentos de protocolos experimentales con base en el $PU_{documentos_pe}$.

En Tabla VI se ilustra el Top-N de experimentos recomendados al usuario con base en la preferencia por un documento del P.E.

TABLA VI
TOP-N EXPERIMENTOS CON BASE EN $PU_{DOCUMENTOS_PE}$

ID del experimento	Basado en protocolo	Porcentaje de similitud
2	LMS-PCJIC-03	98,9297%
35	LMS-PCJIC-03	98,9297%
37	LMS-PCJIC-03	98,9297%
4	LMS-PCJIC-05	98,1886%
79	LMS-PCJIC-05	98,1886%
81	LMS-PCJIC-05	98,1886%
6	LMS-PCJIC-04	98,0548%
58	LMS-PCJIC-04	98,0548%
59	LMS-PCJIC-04	98,0548%
5	LMS-PCJIC-06	97,6579%

Esta prueba se considera satisfactoria de acuerdo con los coeficientes de similitud del P.U (muy cercanos al 100% en su coeficiente de similitud) y la lógica implementada en el componente de contextualización de la recomendación ya que, este componente ofrece al usuario una lista de experimentos en donde convergen los elementos de C.U con mayor coeficiente de similitud con el perfil del usuario.

B. Prueba 2: preferencia por una actividad del flujo de trabajo experimental

En Fig. 11 se ilustran los coeficientes de similitud del perfil del usuario $PU_{procedimientos}$, con todas las actividades individuales del flujo de trabajo experimental registradas en los

experimentos correspondientes a los casos de prueba. En la parte superior de la figura y encerrados en un círculo de color rojo, se encuentran los mejores coeficientes de similitud que corresponden a las actividades individuales de flujo de trabajo experimental que se encuentran registradas con un ID en el rango $1 \leq ID \leq 400$ (aproximadamente). Parte de estos ID son los que se desean ver en el $Top-N_{procedimientos}$.

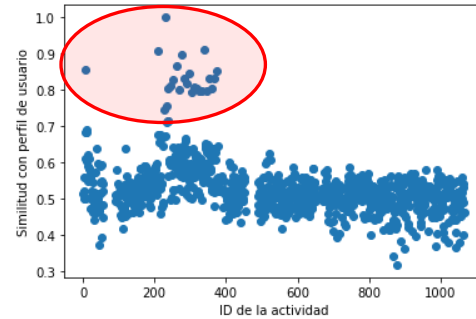


Fig. 11. Coeficientes de similitud de actividades individuales del flujo de trabajo experimental con base en $PU_{procedimientos}$.

En Tabla VII se ilustra el Top-N de experimentos recomendados al usuario con base en la preferencia por una actividad individual del flujo de trabajo experimental.

TABLA VII
TOP-N EXPERIMENTOS CON BASE EN $PU_{PROCEDIMIENTOS}$

ID del experimento	ID de la actividad similar	Porcentaje de similitud
52	339	90,9991%
35	211	90,6756%
43	276	89,6509%
41	262	86,7174%
2	7	85,7212%
57	374	85,3777%
46	297	84,6245%
54	353	83,2935%
56	367	83,2762%
44	283	83,0371%

La prueba también se considera satisfactoria de acuerdo con los mismos criterios de la prueba anterior. Los resultados se consideran relevantes para el usuario.

C. Prueba 3: preferencia por una segunda actividad del flujo de trabajo experimental

Es importante mencionar que, en este caso, el P.U ya incluye dos preferencias en función de las actividades individuales del flujo de trabajo experimental. En Tabla VIII se ilustra el Top-N de experimentos recomendados al usuario con base en este P.U.

TABLA VIII
TOP-N EXPERIMENTOS CON BASE EN SEGUNDA PREFERENCIA $PU_{PROCEDIMIENTOS}$

ID del experimento	ID de la actividad similar	Porcentaje de similitud
88	746	83,3479%
94	794	82,1527%
2	7	80,7495%
52	339	80,4719%
43	276	79,9883%
83	706	79,9034%
90	762	79,8695%
98	826	79,7185%
100	842	79,7175%
96	810	79,4933%

De igual manera la prueba se considera satisfactoria. Aunque los coeficientes de similitud no están sobre el 90%, un porcentaje de similitud del 80% o cercano, sigue siendo un buen resultado ya que el usuario considera los resultados relevantes y relacionados con sus preferencias de perfil de usuario.

D. Prueba 4: preferencia por un recurso de la categoría de materiales y equipos

En la Tabla IX se ilustra el Top-N de experimentos recomendados al usuario con base en la preferencia por un recurso de la categoría de materiales y equipos.

TABLA IX
TOP-N EXPERIMENTOS RECOMENDADOS AL USUARIO CON BASE EN
PU_{RECURSOS_ME}

ID del experimento	Porcentaje similitud	Razón de la recomendación
71	100%	Exp. incluye el mismo recurso
103	95,7472%	Exp. incluye un recurso similar
104	92,3174%	Exp. incluye un recurso similar
65	95,7472%	Exp. incluye un recurso similar
69	100%	Exp. incluye el mismo recurso
68	100%	Exp. incluye el mismo recurso
58	100%	Exp. incluye el mismo recurso
66	100%	Exp. incluye el mismo recurso
113	92,3174%	Exp. incluye un recurso similar
90	92,3174%	Exp. incluye un recurso similar

Esta prueba también se considera satisfactoria de acuerdo con los criterios de las pruebas anteriores. Nótese que los resultados ofrecidos al usuario están considerando los experimentos donde se presenta mayor coeficiente de similitud ya que el mismo recurso puede ser usado en diferentes ensayos de laboratorio.

E. Prueba 5: preferencia por un recurso de la categoría de ecuaciones.

En la siguiente prueba se indica la preferencia por un recurso de tipo ecuación, en particular sobre la ecuación del volumen del cilindro de la expresión 2 que se usa en el protocolo experimental “Guía-LMS-PCJIC-03” ya que en los experimentos asociados a este P.E, unos (13 experimentos) registran la ecuación de la expresión 2 y otros (10 experimentos) la ecuación de la expresión 3.

$$V = \pi \frac{D_2}{4} \times h \quad (2)$$

$$V = \pi \times r^2 \times h \quad (3)$$

En la siguiente tabla se ilustra el Top-N de experimentos recomendados al usuario con base en este P.U

TABLA X
TOP-N EXPERIMENTOS RECOMENDADOS AL USUARIO CON BASE EN
PU_{ECUACIONES}

ID del recurso similar	Porcentaje de similitud	Razón de la recomendación
21	100%	Exp. incluye ecuación de preferencia
2291	93,0629%	Exp. incluye ecuación (3)
21	100%	Exp. incluye ecuación de preferencia
2291	93,0629%	Exp. incluye ecuación (3)
73	27,8265%	Similitud por nombre y descripción
86	27,1655 %	Similitud por nombre y descripción
9	25,5249%	Similitud por nombre y descripción
74	11,8615%	Similitud por nombre y descripción
31	10,7431%	Similitud por nombre y descripción
87	7,2223%	Similitud por nombre y descripción

La prueba se considera parcialmente satisfactoria ya que como el resto de las ecuaciones en el repositorio comparten valores similares para los atributos nombre y descripción, aparecen en la recomendación. Si se consideran relevantes aquellos experimentos que dependen del protocolo LMS-PCJIC-03 que es donde el usuario seleccionó como preferencia la ecuación de la expresión 2, solo el 40% de los resultados obtenidos son relevantes.

F. Fortalezas del modelo

- El modelo de representación permite formalizar el documento del P.E con todas sus secciones y el contenido de estas, permitiendo una representación flexible y cercana al lenguaje natural.
- El modelo de representación permite formalizar el experimento dando existencia de manera detallada a las actividades del flujo de trabajo experimental, los recursos, los autores y los terceros que participan en la actividad experimental del dominio de ingeniería civil. Además, permite representar la existencia de un experimento asociado al documento del P.E que lo define.
- El modelo de recomendación responde al mecanismo de calificación unaria de preferencias explícitas del usuario por los elementos del C.U.
- La lógica que se implementa en el componente de la contextualización de la recomendación obtiene las instancias de los P.Es de cada Top-N en donde se usan los elementos del P.U que configuran sus preferencias de C.U.
- Los coeficientes de similitud calculados en cada una de las pruebas son coherentes con el P.U. La estrategia basada en contenido funciona para calcular las recomendaciones en función de los elementos del C.U.

G. Debilidades del modelo

- La estrategia basada en contenido para calcular la similitud de ecuaciones no parece ser la más adecuada ya que la ecuación como concepto va más allá de sus atributos textuales. Por ejemplo, en el caso de las ecuaciones de las expresiones (2) y (3), ambas se refieren al mismo concepto: volumen de un cilindro, sin embargo, su representación textual varía, lo que hace que el cálculo de similitud en función únicamente de sus atributos textuales no sea el más adecuado.

VII. CONCLUSIONES

Con respecto al problema de la representación formal del P.E en el dominio de ingeniería civil: (a) se propone un modelo flexible de representación que permite unificar el documento del P.E y el experimento con todos los elementos que intervienen en él. (b) Parte de los trabajos previos adelantados en las ontologías, particularmente parte de los aportes en la representación de P.E de la ontología SMART Protocols [4]. (c) Se propone un modelo de representación para un dominio poco explorado como lo es el dominio de ingeniería civil. (d) El modelo permite explotar las descripciones en lenguaje natural.

Con respecto al contexto de uso (C.U): (a) Hace referencia a la usabilidad de productos y en este trabajo el producto es el P.E. producto que con la presencia de ciertos elementos en su formalización lo hacen más o menos usable de acuerdo con

[62]. (b) Se caracteriza el C.U a partir de los elementos del experimento.

Con respecto al modelo de recomendación: (a) Se propone un modelo de recomendación que aprovecha las descripciones textuales que permite el modelo de representación. (b) Se explotan los elementos del C.U como preferencias del usuario. (d) Se está proponiendo trabajar con un objeto complejo como el P.E donde se considera la riqueza y complejidad de las entidades que convergen en él, particularmente en un experimento puntual. Destacar estas entidades individuales es importante es importantes en el proceso de recomendación adaptando, además, esta visión del experimento nos acerca al concepto Objeto de Investigación [64]. (e) Se propone un modelo de recomendación para este tipo de producto científico, modelo que aún no es visible en los repositorios que destacan la utilidad de este artefacto.

Con respecto a la validación del modelo se consideran satisfactorias las pruebas realizadas sin embargo, se sugiere como trabajo futuro: (a) mejorar las recomendaciones con relación a ecuaciones, (b) realizar pruebas en entorno de producción con usuarios, (c) mejorar la integración de resultados finales (contextualización de la recomendación) a partir de los vectores de coeficientes de similitud, (d) ampliar la expresividad del flujo de trabajo experimental más allá de actividades secuencialmente ordenadas, (e) probar con otras técnicas de procesamiento del lenguaje natural, (f) trabajar en aspectos de arquitectura con respecto al rendimiento: cálculo y recálculo de las recomendaciones.

REFERENCIAS

- [1] M. Assante, "Science 2.0 Repositories," UNIVERSITÀ DI PISA, 2015
- [2] S. Crabu, "Give us a protocol and we will rise a lab. The shaping of the infra-structuring objects," *Inf. Infrastructure(s) boundaries, Context. Ecol.*, pp. 120–165, 2014.
- [3] O. Giraldo, A. Garcia, and O. Corcho, "A guideline for reporting experimental protocols in life sciences," *PeerJ*, vol. 6, p. e4795, 2018, doi: 10.7717/peerj.4795.
- [4] O. Giraldo, A. García, F. López, and O. Corcho, "Using semantics for representing experimental protocols," *J. Biomed. Semantics*, vol. 8, no. 1, 2017, doi: 10.1186/s13326-017-0160-y.
- [5] L. Teytelman, A. Stoliartchouk, L. Kindler, and B. L. Hurwitz, "Protocols.io: Virtual Communities for Protocol Development and Discussion," *PLoS Biol.*, 2016, doi: 10.1371/journal.pbio.1002538.
- [6] D. De Roure *et al.*, "Towards open science: The myExperiment approach," *Concurr. Comput. Pract. Exp.*, 2010, doi: 10.1002/cpe.1601.
- [7] R. Palma, P. Holubowicz, O. Corcho, J. M. Gómez-Pérez, and C. Mazurek, "Rohub - A digital library of research objects supporting scientists towards reproducible science," *Commun. Comput. Inf. Sci.*, 2014, doi: 10.1007/978-3-319-12024-9_9.
- [8] L. N. Soldatova *et al.*, "EXACT2: the semantics of biomedical protocols," *BMC Bioinformatics*, vol. 15 Suppl 1, no. Suppl 14, p. S5, 2014, doi: 10.1186/1471-2105-15-S14-S5.
- [9] D. Jannach, M. Zanker, A. Felfernig, and F. Gerhard, "Content-Based Recommendation," in *Recommender Systems an Introduction*, Cambridge University Press, 2012.
- [10] D. Asanov, "Algorithms and Methods in Recommender Systems," *Other Conf.*, 2011, doi: 10.1146/annurev-psych-113011-143823.
- [11] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based Recommender Systems: State of the Art and Trends," in *Recommender Systems Handbook*, Springer Science+Business Media, 2011, pp. 73–105.
- [12] ISO, "ISO 9241-11," *Ergonomic requirements for office work with visual display terminals (VDIs) -- Part 11: Guidance on usability*, 1998. .
- [13] T. Kunert, "Interactive TV Applications and Their Context of Use," in *User-centered Interaction Design Patterns for Interactive Digital Television Applications*, London: Springer-Verlag, 2009, pp. 19–47.
- [14] M. Maguire, "Context of use within usability activities," in *International Journal of Human Computer Studies*, 2001, vol. 55, no. 4, pp. 453–483, doi: 10.1006/ijhc.2001.0486.
- [15] M. K. Pinheiro, "Adaptation Contextuelle et Personnalisée de l'Information de Conscience de Groupe au sein des Systèmes d'Information Coopératifs," Université Joseph-Fourier - Grenoble I, 2006.
- [16] A. Carrillo-Ramos, M. Villanova-Oliver, J. Gensel, and H. Martin, "Profiling nomadic users considering preferences and context of use," in *ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS 2007: OTM 2007 WORKSHOPS, PT 1, PROCEEDINGS*, 2007, vol. 4805, pp. 457–466.
- [17] J. M. C. Fonseca *et al.*, "Model-Based UI XG Final Report," *W3C Incubator Gr. Report*, May, p. 32, 2010.
- [18] D. Allemang and J. Hendler, "Semantic Web for the Working Ontologist: Effective modeling in RDFS and OWL," *Semantic Web for the Working Ontologist*. Elsevier, pp. 2–12, 2011.
- [19] R. Studer, V. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods," *Data Knowl. Eng.*, vol. 25, no. 1–2, pp. 161–197, 1998, doi: 10.1016/S0169-023X(97)00056-6.
- [20] L. N. Soldatova and R. D. King, "An ontology of scientific experiments," *J. R. Soc. Interface*, vol. 3, no. June, pp. 795–803, 2006, doi: 10.1098/rsif.2006.0134.
- [21] A. Maccagnan *et al.*, "Combining ontologies and workflows to design formal protocols for biological laboratories.," *Autom. Exp.*, vol. 2, p. 3, 2010, doi: 10.1186/1759-4499-2-3.
- [22] A. Bandrowski *et al.*, "The Ontology for Biomedical Investigations," *PLoS One*, vol. 11, no. 4, p. e0154556, 2016, doi: 10.1371/journal.pone.0154556.
- [23] O. Giraldo, "SMART Protocols." 2017.
- [24] K. Belhajjame *et al.*, "Using a suite of ontologies for preserving workflow-centric research objects," *J. Web Semant.*, vol. 32, pp. 16–42, 2015, doi: 10.1016/j.websem.2015.01.003.
- [25] I. Torres, J. Guzmán, and F. Moreno, "Semántica para un sistema de recuperación de protocolos experimentales en ingeniería civil," *Coloq. Investig. Multidiscipl.*, vol. 5, no. 2, p. 753, 2017.
- [26] S. Ajmani, H. Ghosh, A. Mallik, and S. Chaudhury, "An ontology based personalized garment recommendation system," in *Proceedings - 2013 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW 2013*, 2013, vol. 3, pp. 17–20, doi: 10.1109/WI-IATW.2013.143.
- [27] M. Asaduzzaman, C. K. Roy, S. Monir, and K. A. Schneider, "Exploring API method parameter recommendations," in *2015 IEEE 31st International Conference on Software Maintenance and Evolution, ICSME 2015 - Proceedings*, 2015, pp. 271–280, doi: 10.1109/ICSM.2015.7332473.
- [28] C. Bologna, A. C. De Rosa, A. De Vivo, M. Gaeta, G. Sansonetti, and V. Viserta, "Personality-based recommendation in E-commerce," in *CEUR Workshop Proceedings*, 2013, vol. 997.
- [29] D. D'Agostino, F. Gasparetti, A. Micarelli, and G. Sansonetti, "A social context-aware recommender of itineraries between relevant points of interest," in *Communications in Computer and Information Science*, 2016, vol. 618, pp. 354–359, doi: 10.1007/978-3-319-40542-1_58.
- [30] Y. García J., M. García Y.S., and S. M. González B., "Semantic Characterization of Context of Use and Contents for User-Centric Media Retrieval," *User Centric Media. UCMEDIA 2010*, vol. 60, pp. 20–25, 2012, doi: 978-3-642-35145-7_3.
- [31] H. Lee, Y. Choi, and Y.-J. Kim, "An adaptive user interface based on spatiotemporal structure learning," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 118–124, 2011, doi: 10.1109/MCOM.2011.5783996.
- [32] D. Lichtnow, I. Gasparini, A. Bouzeghoub, J. P. M. De Oliveira, and M. S. Pimenta, "Recommendation of learning material through students' collaboration and user modeling in an adaptive e-learning environment," *Studies In Computational Intelligence*, vol. 350. Springer-Verlag, Berlin Heidelberg, pp. 257–278, 2011, doi: 10.1007/978-3-642-19814-4_12.
- [33] K. Niemann, "Automatic tagging of learning objects based on their usage in web portals," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9307, pp. 240–253, doi: 10.1007/978-3-319-24258-3_18.

- [34] M. A. Orthmann *et al.*, “Usage-based clustering of learning objects for recommendation,” in *Proceedings of the 2011 11th IEEE International Conference on Advanced Learning Technologies, ICALT 2011*, 2011, pp. 553–557, doi: 10.1109/ICALT.2011.169.
- [35] H. Väättäjä, “Characterizing the context of use in mobile work,” in *IFIP Advances in Information and Communication Technology*, 2015, vol. 468, pp. 97–113, doi: 10.1007/978-3-319-27048-7_7.
- [36] Z. Xu, F. Zhang, W. Wang, H. Liu, and X. Kong, “Exploiting Trust and Usage Context for Cross-Domain Recommendation,” *IEEE Access*, vol. 4, pp. 2398–2407, 2016, doi: 10.1109/ACCESS.2016.2566658.
- [37] D. Zakrzewska, “Cluster analysis in personalized e-learning systems,” *Stud. Comput. Intell.*, vol. 252, pp. 229–250, 2009, doi: 10.1007/978-3-642-04170-9_10.
- [38] D. Zakrzewska, “Building context-aware group recommendations in E-learning systems,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6922 LNAI, no. PART 1, pp. 132–141, doi: 10.1007/978-3-642-23935-9_13.
- [39] Nature Publishing Group, “Protocol Exchange,” 2018. [Online]. Available: <https://www.nature.com/protocolexchange/protocols/>. [Accessed: 03-Oct-2018].
- [40] OpenWetWare Contributors, “OpenWetWare,” 2017. [Online]. Available: <https://openwetware.org/>. [Accessed: 03-Oct-2018].
- [41] Springer Nature, “Springer Nature Experiments,” 2018. [Online]. Available: <https://experiments.springernature.com/>. [Accessed: 03-Oct-2018].
- [42] K. Wolstencroft *et al.*, “FAIRDOMHub: A repository and collaboration environment for sharing systems biology research,” *Nucleic Acids Res.*, 2017, doi: 10.1093/nar/gkx1032.
- [43] Benchling Inc., “Benchling,” 2018. [Online]. Available: <https://benchling.com/>. [Accessed: 03-Oct-2018].
- [44] SciNote, “SciNote,” 2018. [Online]. Available: <https://scinote.net/>. [Accessed: 03-Oct-2018].
- [45] LabArchives LLC., “LabArchives - Notebook,” 2018. [Online]. Available: <https://www.labarchives.com/>. [Accessed: 03-Oct-2018].
- [46] Opentrons, “Opentrons,” 2018. [Online]. Available: <https://opentrons.com/>. [Accessed: 03-Oct-2018].
- [47] The CellML Project, “cellML,” 2018. [Online]. Available: <https://www.cellml.org/>. [Accessed: 03-Oct-2018].
- [48] Treat-NMD Neuromuscular Network, “Experimental Protocols for DMD animal models,” 2018. [Online]. Available: <http://www.treat-nmd.eu/research/preclinical/dmd-sops/>. [Accessed: 03-Nov-2018].
- [49] J. Spidlen, K. Breuer, C. Rosenberg, N. Kotecha, and R. R. Brinkman, “FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications,” *Cytom. Part A*, vol. 81A, no. 9, pp. 727–731, 2012, doi: 10.1002/cyto.a.22106.
- [50] UC San Diego, “The Metabolomics Workbench,” 2018. [Online]. Available: <http://www.metabolomicsworkbench.org/>. [Accessed: 03-Oct-2018].
- [51] Swiss Institute of Bioinformatics, “Swiss-Model,” 2018. [Online]. Available: <https://swissmodel.expasy.org/repository/>. [Accessed: 03-Oct-2018].
- [52] Utrecht University, “Search Repository,” 2018. [Online]. Available: <https://uiloos-labs.wp.hum.uu.nl/resources/search-resources/>. [Accessed: 03-Oct-2018].
- [53] Bio-protocol LLC., “Bio-Protocol,” 2018. [Online]. Available: <https://bio-protocol.org/>. [Accessed: 03-Oct-2018].
- [54] BMC, “Biological Procedures Online,” 2018. [Online]. Available: <https://biologicalproceduresonline.biomedcentral.com/>. [Accessed: 03-Oct-2018].
- [55] Dryad Digital Repository, “Dryad,” 2016. [Online]. Available: <http://datadryad.org/>. [Accessed: 03-Oct-2018].
- [56] P. J. Hane, “Sharing Research Data—New figshare For Institutions,” *Against The Grain*, 2013. [Online]. Available: <http://www.against-the-grain.com/2013/09/sharing-research-data-new-figshare-for-institutions-2/>. [Accessed: 03-Oct-2018].
- [57] JoVE, “JoVE - Video Journal,” 2018. [Online]. Available: <https://www.jove.com/>. [Accessed: 03-Oct-2018].
- [58] K. Belhajjame *et al.*, “Workflow-centric research objects: First class citizens in scholarly discourse,” in *CEUR Workshop Proceedings*, 2012, vol. 903, pp. 1–12, doi: 10.1.1.383.203.
- [59] ASTM Standard D2216-10, “Standard Test Methods for Laboratory Determination of Water (Moisture) Content of Soil and Rock by Mass,” 2010.
- [60] V. E. Meza Ochoa, *Guía de Laboratorio de Mecánica de Suelos*. Medellín: Politécnico Colombiano Jaime Isaza Cadavid, 2013.
- [61] Y. Zheng, “Tutorial: Context-Awareness In Information Retrieval and Recommender Systems,” in *The 16th IEEE/WIC/ACM Conference on Web Intelligence*, 2016.
- [62] O. Giraldo, A. Garcia, and O. Corcho, “A guideline for reporting experimental protocols in life sciences,” *PeerJ*, vol. 6, p. e4795, 2018, doi: 10.7717/peerj.4795.
- [63] K. M. Ting, “Precision and Recall,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, p. 781.
- [64] S. Bechhofer, D. De Roure, M. Gamble, C. Goble, and I. Buchan, “Research Objects: Towards Exchange and Reuse of Digital Knowledge,” *Nat. Preced.*, 2010, doi: 10.1038/npre.2010.4626.

Juan Felipe Muñoz-Fernández es Ingeniero Informático egresado del Politécnico Colombiano Jaime Isaza Cadavid . Se graduó como Especialista en Sistemas de la Universidad Nacional de Colombia, sede Medellín y como MSc. en Ingeniería – Ingeniería de Sistemas, título que también obtuvo en la Universidad Nacional de Colombia, sede Medellín. Actualmente es docente ocasional adscrito al Departamento de Ciencias de la Computación y de la Decisión de la Universidad Nacional de Colombia, sede Medellín.
ORCID: <https://orcid.org/0000-0003-4810-8873>.

Jaime Alberto Guzmán-Luna es Ingeniero Civil egresado de la Universidad Nacional de Colombia, sede Medellín. Obtuvo su título de MSc. en Ingeniería – Ingeniería de Sistemas en la misma universidad y es Ph.D. en Ingeniería, también de la Universidad Nacional de Colombia, sede Medellín. Actualmente lidera el grupo de investigación Sistemas Inteligentes Web (SINTELWEB) y es docente titular adscrito a al Departamento de Ciencias de la Computación y de la Decisión de la Universidad Nacional de Colombia, sede Medellín. ORCID: <https://orcid.org/0000-0003-4737-1119>