

PROYECTO DE INVESTIGACIÓN INTERNO PII-18-01
“Semántica embebida con JSON-LD para generar información enriquecida en resultados de búsqueda de Recursos Educativos Abiertos”

En la ciudad de Quito D.M., a los dos días del mes de septiembre del año dos mil veinte, comparecen a la celebración de la presente Acta de Finalización del Proyecto de Investigación Interno **PII-18-01 “Semántica embebida con JSON-LD para generar información enriquecida en resultados de búsqueda de Recursos Educativos Abiertos”**, por una parte la **Ph.D. Alexandra Patricia Alvarado Cevallos** en calidad de **Vicerrectora de Investigación, Innovación y Vinculación** de la Escuela Politécnica Nacional, y por otra la **Ph.D. Rosa del Carmen Navarrete Rueda** en calidad de **Directora del Proyecto de Investigación Interno PII-18-01**, al tenor de lo siguiente:

1. ANTECEDENTES:

- a) El 19 de junio de 2018, el Consejo de Investigación y Proyección Social mediante Resolución R088/18, aprueba el Cronograma para la presentación de Propuestas de Proyectos de Investigación – Convocatoria 2018, y mediante Resolución R090/18 del 10 de julio de 2018 se aprueba el cronograma modificado. El 31 de julio de 2018, el Consejo de Investigación y Proyección Social, mediante Resolución R096/18 aprueba una nueva modificación del cronograma para la presentación de Proyectos de Investigación, y mediante Resolución R160/18 del 13 de noviembre de 2018, se aprueba el cronograma final para la presentación de Proyectos de Investigación – Convocatoria 2018.
- b) El 5 de febrero de 2019, al amparo de lo dispuesto por Consejo de Investigación y Proyección Social, mediante Resolución RCIPS-028-2019, se aprobó el “Informe Final de la Convocatoria 2018”, donde se mostraron los resultados y los proyectos aprobados de dicha convocatoria. Entre los proyectos aprobados se encuentra el proyecto Interno denominado “*Semántica embebida con JSON-LD para generar información enriquecida en resultados de búsqueda de Recursos Educativos Abiertos*”, presentado por la Ph.D. Rosa Navarrete.
- c) Mediante Memorando EPN-VIPS-2019-0436-M del 19 de marzo de 2019, el Vicerrectorado de Investigación y Proyección Social informa a los directores de los proyectos internos 2018 que el inicio de los proyectos es el 1 de abril de 2019.

2. DATOS GENERALES DEL PROYECTO:

| | |
|----------------------------------|---|
| Código de Proyecto | <i>PII-18-01</i> |
| Nombre del Proyecto | <i>Semántica embebida con JSON-LD para generar información enriquecida en resultados de búsqueda de Recursos Educativos Abiertos</i> |
| Directora del Proyecto | <i>NAVARRETE RUEDA ROSA DEL CARMEN</i> |
| Codirector del Proyecto | <i>MONTENEGRO ARMAS CARLOS ESTALESMIT WILLAM</i> |
| Colaboradora del Proyecto | <i>RECALDE CERDA LORENA KATHERINE</i> |
| Departamento | <i>Informática y Ciencias de la Computación (DICC)</i> |
| Líneas de Investigación | <i>Sistemas de información</i> |
| Objetivo | <i>Desarrollar un esquema de semántica embebida con JSON-LD y vocabulario estandarizado para enriquecer información de resultados de búsqueda de Recursos Educativos Abiertos</i> |
| Duración del Proyecto | <ul style="list-style-type: none"> • <i>Inicio: 1 de abril del 2019</i> • <i>Fin planificado: 31 de marzo del 2020</i> • <i>Prórroga técnica: 3 meses, hasta el 30 de junio de 2020</i> • <i>Duración total: 15 meses</i> |

| | |
|----------------------------------|--|
| Entrega del Informe Final | <i>20 de agosto del 2020</i> |
| Presupuesto asignado | <i>\$ 4.999,68 USD (cuatro mil novecientos noventa y nueve dólares americanos, con 68/100)</i> |
| Presupuesto ejecutado | <i>\$ 3.390,90 USD (tres mil trescientos noventa dólares americanos, con 90/100)</i> |

3. INFORME FINAL:

Mediante Memorando EPN-PII-18-01-2020-0008-M del 20 de agosto de 2020, la Ph.D. Rosa Navarrete, Directora del Proyecto PII-18-01, presenta el Informe Final del Proyecto Interno, esta documentación es revisada por la Dirección de Investigación, se anexa y forma parte integrante del Acta de Finalización, cuyas conclusiones y productos generados son:

CONCLUSIONES:

El World Wide Web Consortium (W3C), organismo rector de la Web, produce estándares, tecnologías y recomendaciones que proyectan el crecimiento de la Web a largo plazo. La propuesta es que tales productos sean utilizados en forma global y aplicados en todos los dominios del conocimiento y del quehacer humano. Sin embargo, quienes publican contenidos en la Web no tienen necesariamente el conocimiento para hacer uso de tales estándares y tecnologías.

En este proyecto de investigación se obtuvo una revisión de la extensión en el uso de anotaciones semánticas embebidas, marcadas con JSON-LD, en el contenido web de Recursos Educativos Abiertos.

Se empleó como base el último corpus web liberado al dominio público en la fecha de investigación, obtenido del Common Crawl. Este corpus web es extraído mediante un proceso de rastreo de la Web y contiene millones de páginas web, lo que lo hace un corpus fiable para análisis cuantitativo de los estándares utilizados en tales páginas.

Los resultados permitieron establecer en qué medida dicha tecnología estaba siendo utilizada y proponer una herramienta que ponga al alcance de los usuarios finales una aplicación web para generar scripts de marcado de modo que los creadores de recursos educativos, sin requerir conocimiento de tecnología, sino únicamente del contexto educativo, lo puedan incluir en su contenido web para incrementar la visibilidad futura de tal contenido.

- a. En el objetivo 1, "Analizar la extensión de uso de la especificación JSON-LD mediante el análisis cuantitativo de corpus de rastreo web", se esperaba encontrar un uso intensivo del formato JSON-LD, debido a que alcanzó el status de recomendación de la W3C desde enero de 2014. Los resultados demostraron que el despliegue de esta tecnología es aún incipiente.
- b. En el objetivo 2, "Identificar los descriptores del vocabulario Schema relevantes al contexto educativo, de mayor uso en los corpus de rastreo web", se buscaba determinar cuáles eran los descriptores más utilizados para etiquetar el contenido educativo y de aquellos, cuáles provenían del vocabulario de Schema. Se encontró que los descriptores no eran todos los esperados sino únicamente un grupo de ellos. Adicionalmente, algunos de esos descriptores no estaban correctamente utilizados, lo que provocaría que el motor de búsqueda ignore el marcado. Se obtuvo un análisis cuantitativo de la ocurrencia de 26 propiedades, de clase superior y términos básicos asociados al dominio educativo. Dichas propiedades provenían del vocabulario de Schema y derivados de especificaciones como LRMI.
- c. En el objetivo 3, "Definir una recomendación para el marcado semántico embebido de REA utilizando JSON-LD y vocabulario seleccionado de Schema o complementario a Schema", se analizó la utilidad de la información que debería ser proporcionada en el marcado embebido de los recursos educativos abiertos, para que al ser procesado por un motor de búsqueda como Google, permita obtener información enriquecida en los resultados de búsqueda, que guíen

efectivamente al usuario en sus decisiones de exploración. Bajo esta perspectiva se realizó una recomendación de las propiedades comunes para todos los tipos de recurso educativo y aquellas particulares a cada recurso.

- d. En el objetivo 4, "Desarrollar un prototipo generador automático de scripts para el marcado de REA con JSON-LD", se desarrolló una aplicación web, como herramienta en línea, que permite la generación automática del script de marcado en formato JSON-LD. Para utilizar esta herramienta, la persona interesada únicamente debe conocer la información concerniente al contexto del recurso educativo que está publicando en la web. Esta herramienta guía paso a paso para la producción del contenido del marcado, hasta obtener un script JSON-LD validado. Este es un fragmento que puede ser copiado y pegado en la página web del recurso educativo. Se propone que su uso permita acercar a los usuarios finales que puedan utilizar de forma simple esta herramienta de marcado semántico embebido.

Como conclusión puede señalarse que este proyecto ha contribuido con una referencia cuantitativa del uso del formato de marcado semántico embebido JSON-LD. Es importante, ya que, al ser una recomendación de la W3C, se esperaba un uso intensivo y extensivo de ese formato; sin embargo, se demostró que, en el campo educativo, aún no es un formato difundido.

La propuesta de marcado con JSON-LD y su simplificación a través de una herramienta automática para generación, planteada en este proyecto, contribuye en mejorar y simplificar la capacidad de los usuarios para encontrar, reutilizar, crear y compartir Recursos Educativos Abiertos. Al simplificar este proceso, ofreciendo información semántica en los resultados de búsqueda, se utilizará efectivamente el potencial que ofrecen los motores de búsqueda actuales en la construcción de conocimiento a partir de la semántica embebida. El potencial uso de los REA ha sido reconocido por UNESCO como base para la consecución del Objetivo de educación de calidad, como parte de los Objetivos de desarrollo del milenio propuesto por las Naciones Unidas en la Agenda 2030 y que ha sido adoptada por gobiernos en todo el mundo, incluido el gobierno de Ecuador, como política pública.

PRODUCTOS:

- Artículo en congreso por pares: "*Analyzing embedded semantic with JSON-LD and microdata for educational resources in large scale web datasets*" (Scopus); Navarrete R., Recalde L., Montenegro C., Luján Mora S; *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*; ISBN: 978-172815584-5; DOI: 10.1109/CSCI49370.2019.00214; diciembre 2019.
- Artículo: "*Systematic Mapping on Embedded Semantic Markup validated with Data Mining techniques*"; Rosa Navarrete, Carlos Montenegro, Lorena Recalde; *Advances in Artificial Intelligence, Software and Systems Engineering (Springer)*; ISBN: 978-3-319-94229-2; DOI: 10.1007/978-3-030-51328-3_53; julio 2020.
- Presentación en conferencia: "*Analyzing Embedded Semantic with JSON-LD and Microdata for Educational Resources in Large Scale Web Datasets*"; Navarrete Rosa; 2019 International Conference on Computational Science and Computational Intelligence (CSCI'19); Las Vegas – Nevada, Estados Unidos; diciembre 2019.
- Presentación en conferencia virtual: "*Systematic Mapping on Embedded Semantic Markup validated with Data Mining techniques*"; Navarrete Rosa; 11th International Conference on Applied Human Factors and Ergonomics (AHFE 2020); San Diego – California, Estados Unidos; julio 2020.

4. LIQUIDACIÓN ECONÓMICA:

El monto asignado al Proyecto Interno PII-18-01 fue de \$ 4.999,68 USD (cuatro mil novecientos noventa y nueve dólares americanos, con 68/100), y se ejecutaron \$ 3.390,90 USD (tres mil trescientos noventa dólares americanos, con 90/100), conforme al detalle emitido por la Unidad de Gestión de

Investigación y Proyección Social del Vicerrectorado de Investigación, Innovación y Vinculación, que se adjunta a la presente Acta y forma parte integrante de la misma.

5. FINALIZACIÓN:

Con la presente Acta se declara finalizado y cerrado el Proyecto Interno PII-18-01 "*Semántica embebida con JSON-LD para generar información enriquecida en resultados de búsqueda de Recursos Educativos Abiertos*".

Para constancia de lo ejecutado y por estar de acuerdo con el contenido de la presente Acta, las partes libre y voluntariamente suscriben la misma, en tres ejemplares de igual contenido, tenor y valor legal.

Dado en la ciudad de Quito, D.M. a los dos días del mes de septiembre del año dos mil veinte.



Firmado electrónicamente por:
**ROSA DEL CARMEN
NAVARRETE RUEDA**

Ph.D. Alexandra Alvarado
**Vicerrectora de Investigación,
Innovación y Vinculación**

Ph.D. Rosa Navarrete
**Directora del Proyecto
PII-18-01**

sp/cr