



Diseño y generación de un repositorio de información de salud para la ciencia de datos

Design and creation of a health information repository for data science

MARÍA LUISA ÁVILA MORALES^a, VÍCTOR MORALES ROCHA^{a*}, BENITO A. PONCE RODRÍGUEZ^a, RAMÓN PARRA LOERA^a
^aMaestría en Cómputo Aplicado, Departamento de Ingeniería Eléctrica y Computación, Instituto de Ingeniería y Tecnología, Universidad Autónoma de Ciudad Juárez, México.

*Autor de correspondencia. Correo electrónico: victor.morales@uacj.mx

No. de resumen 4CP22-32	Formato Ponencia
Evento 4.º Coloquio de Posgrados del IIT	Presentador María Luisa Ávila Morales
Tema Cómputo Aplicado	Estatus Estudio en curso
Fecha de la presentación Noviembre 22, 2022	

Resumen

El objetivo de este proyecto es el diseño y generación de un repositorio de información de salud, disponible bajo una licencia de datos abiertos, con el propósito de realizar ciencia de datos. La información que contendrá el repositorio provendrá de Sistemas de Expediente Clínico Electrónico que participen en la plataforma de interoperabilidad MISECE (Módulo de Interoperabilidad para el Sistema de Expediente Clínico Electrónico). El MISECE proveerá los datos del Expediente Clínico Electrónico (ECE) en formato JSON de acuerdo con el estándar FHIR (Fast healthcare Interoperability Resources), un archivo por paciente por entidad participante. Se encuentra en desarrollo y pruebas una aplicación que al recibir el ECE en formato JSON, selecciona los datos apropiados para investigación clínica y epidemiológica, además anonimiza los datos sensibles conforme a las normas mexicanas sobre la privacidad de los datos personales. Estos datos son reservados en una base de datos temporal SQL para continuar con la etapa de transformación de la información. Una meta es el facilitar la extracción de conocimiento del repositorio de datos de salud y satisfacer las políticas de datos abiertos, se planea el uso de una base de datos de grafos y herramientas de Web Semántica de uso gratuito.

Palabras clave: expediente clínico electrónico, datos clínicos, ciencia de datos, datos abiertos, base de datos de grafos.

Abstract

The objective of this project is the design and creation of a repository of health information for data science purpose under an open data license. The repository will hold data from Electronic Health Record (EHR) systems that participate in the interoperability platform MISECE (Módulo de Interoperabilidad para el Sistema de Expediente Clínico Electrónico). MISECE will provide data from EHRs in JSON format according to FHIR standard (Fast healthcare Interoperability Resources), one file per patient per participating entity. An application is being developed and tested which once it receives the EHR in JSON format, selects suitable data for clinical epidemiological research and anonymizes the sensitive data to comply with Mexican regulations regarding personal data privacy. The data is temporarily reserved in an SQL database so it can continue with the following phase, the data transformation. A graph database and freeware Semantic Web tools are in plans to be used to facilitate knowledge extraction from the health data repository and to satisfy open data policies. Graph databases have multiple advantages



over other NoSQL and SQL databases, besides having a relatively constant query time execution they are also easy to model.

Keywords: electronic health record, clinical data, data science, open data, graph database.

Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

Financiamiento

Los autores y beca CONACYT del autor principal.

Conflictos de interés

Los autores declaran que no existe ningún conflicto de interés.