

Noticias del Departamento de Ingeniería de Sistemas

Novedades



- *Artículos presentados en CoNaISI 2022*
- *Tesis dirigida en el departamento*

Artículos en CoNaISI presentados como posters en la categoría de trabajos de estudiantes



Un Proceso de Big Data aplicado a datos de siembra, cosecha, producción y rendimiento del girasol

por MARIANO CONCHILLO,
AGUSTINA BUCCELLA

Este trabajo fue elegido dentro de los 6 mejores para ser presentados (además de póster) como presentación oral.

Resumen. La producción del girasol en nuestro país es una economía regional que involucra a miles de productores, en su mayoría pequeños y medianos, donde gran parte de la actividad económica de las regiones involucradas se mueve en torno al desarrollo de este cultivo. El Ministerio de Agricultura, Ganadería y Pesca ha liberado datos históricos de

siembra, cosecha, producción y rendimiento del girasol relevados en estas regiones, con el objetivo de que la comunidad pueda procesarlos para identificar patrones u otro tipo de comportamientos que puedan ser útiles.

Es por eso, que en este trabajo describimos la aplicación de un proceso de Big Data que toma un gran conjunto de datos y realiza los pasos necesarios para obtener información útil que permita (1) analizar tendencias y relaciones entre las diferentes variables, y (2) realizar predicciones de superficie cosechada, producción y rendimiento. Siguiendo estos dos objetivos, describimos el proceso aplicado junto con los resultados obtenidos.



Un Proceso de Big Data aplicado al transporte con tarjetas SUBE

por FEDERICO SAURIN, JOSÉ LUIS POLO SOLA, AGUSTINA BUCCELLA

Resumen. El transporte público es el medio más utilizado en el mundo y miles de personas dependen de él para poder trasladarse. Es por esto, que su buen funcionamiento, lo que incluye calidad, eficiencia y eficacia tanto en las unidades, viajes y rutas, es fundamental. La utilización de Sistemas de Análisis de Datos o Big Data pueden ser de gran utilidad para analizar y proponer mejoras en este sistema.

Así, en este trabajo describimos la aplicación de un proceso de Big Data que toma un gran conjunto de datos y realiza los pasos necesarios para obtener información útil que permita (1) la identificación de tendencias en el servicio de transporte y (2) la predicción del uso del transporte. Siguiendo estos dos objetivos, describimos el proceso aplicado junto con los resultados obtenidos.

Tesis dirigida en el departamento

El 22 de Diciembre de 2022, el estudiante Luciano Zúñiga defendió su tesis de Licenciatura en Ciencias de la Computación titulada:

Diseño e Implementación de Servicios Estandarizados dentro de



una Línea de Productos de Software aplicada al Dominio de Paleontología

Resumen. La Ingeniería de Software Orientada a Reuso intenta reutilizar componentes de software ya existentes, maximizando el tiempo de desarrollo, la puesta en el mercado y los costos, mejorando así la calidad del producto resultante. Existen diferentes tipos de reuso, en este trabajo vamos a centrar en el reuso orientado a dominios, que se conoce como Ingeniería de Líneas de Producto de Software (ILPS).

En trabajos previos, dentro del proyecto de investigación que se enmarca esta tesis, se ha focalizado en identificar características comunes y variables dentro del dominio geográfico. Se ha analizado el dominio oceanogeográfico para enfocarse en el subdominio de ecología marina, donde se ha desarrollado una Línea de Producto de Software (LPS) que fue reutilizada en diferentes aplicaciones y por diferentes usuarios. Al mismo tiempo, se ha creado una metodología de desarrollo de una LPS y se ha definido una serie de artefactos de software reutilizables.

En esta tesis, a través de un proceso de especialización del dominio geográfico, nos enfocamos en modelar el subdominio de paleontología. Se propone continuar el enfoque definido en trabajos previos para el desarrollo de funcionalidades propias de una LPS en este subdominio, donde extenderemos los artefactos de software previamente creados. Especialmente, se trabajará con la información de estándares que existen para este dominio. Además se obtendrá información de usuarios expertos del Museo de Ciencias Naturales de la Universidad Nacional del Comahue. Las funcionalidades creadas en este trabajo serán luego instanciadas para crear un producto para este museo.

La tesis fue dirigida por la Dra. Agustina Buccella.

Competitividad Digital

Evolución del arte rupestre a través de aprendizaje de máquina!



Trabajando con los propietarios tradicionales de Mimal y Marrku del área del río Wilton, investigadores del sur de Australia han estado utilizando el aprendizaje automático para analizar imágenes de arte rupestre y trazar un mapa de su evolución.

El estudio, dirigido por el arqueólogo de la Universidad de Flinders¹, el Dr. Daryl Wesley, consideró diferentes estilos de arte rupestre de figuras humanas en Arnhem Land etiquetadas como “figuras que corren del norte”, “figuras dinámicas”, “figuras posdinámicas” y “figuras simples con boomerangs” para comprender cómo se relacionan los estilos entre sí. Se utilizó el aprendizaje automático para analizar imágenes de arte rupestre recopiladas en la zona Marrku en 2018 y 2019.

El enfoque utilizó modelos de redes neuronales convolucionales previamente entrenados y publicados y combinaciones de conjuntos de datos que fueron diseñados y entrenados para la clasificación de objetos. Este enfoque permite una clasificación con menos sesgo del estilo, además de hacer uso de redes neuronales sin tener que acceder a un conjunto de datos de arte rupestre de gran tamaño, lo que sería necesario para entrenar un modelo desde cero.

El análisis de la activación del modelo se realizó utilizando la técnica de t-SNE (t-distributed stochastic neighbour embedding), que es un método no lineal para la reducción de

la dimensionalidad. Este método ayuda a dar sentido e interrogar por qué los modelos se han activado de la forma en que lo han hecho para los diferentes puntos de datos.

La cronología del arte rupestre reconstruida utiliza conjuntos de datos existentes de más de 14 millones de fotos diferentes de animales como perros, gatos, lagartijas e insectos hasta objetos como sillas, mesas y tazas. En total, la computadora vio más de 1000 tipos diferentes de objetos y aprendió a distinguirlos con solo mirar fotos de ellos.

La habilidad importante que desarrolló esta computadora fue un modelo matemático que tiene la capacidad de decir qué tan similares son dos imágenes diferentes entre sí. La metodología eliminó un alto grado de interpretación humana individual y posibles sesgos mediante el uso de un enfoque de aprendizaje automático llamado *aprendizaje por transferencia*.

El aprendizaje por transferencia es un análisis en el aprendizaje automático que se enfoca en almacenar el conocimiento obtenido al resolver un problema y aplicarlo a otro problema diferente pero relacionado. Por ejemplo, el conocimiento adquirido al aprender a reconocer automóviles podría aplicarse al tratar de reconocer camiones. Esto permitió que la computadora entendiera cómo cada estilo se relacionaba entre sí directamente, independientemente de los investigadores.

Lo emocionante de la investigación no es que esté encontrando cosas que los humanos han pasado por alto, sino que está replicando los resultados de otros estudios que han utilizado un enfoque más tradicional. Esto demuestra que el enfoque está funcionando y sugiere un potencial emocionante para contribuir a los estudios de arte rupestre en otros lugares.

Si te interesa el tema, puedes ver más del caso de Australia en ² o acceder a información en publicaciones especializadas, como “Artificial

¹<https://www.flinders.edu.au/>

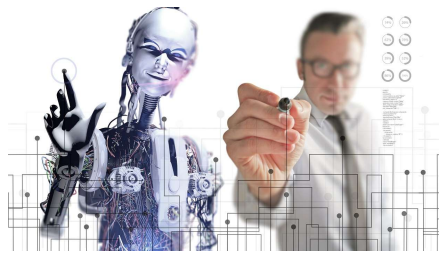
²<https://www.canberratimes.com.au/story/7804586/ai-to-help-unearth-indigenous-rock-art/>



Intelligence, 3D Documentation, and Rock Art—Approaching and Reflecting on the Automation of Identification and Classification of Rock Art Images”³.

Mesa del Arquitecto

Intereses arquitectónicos en el diseño de gemelos digitales



Las organizaciones que adoptan enfoques basados en gemelos digitales (GD) a menudo apuntan a estrategias de recolección de datos tanto reactivas como proactivas. Los GD no sólo ayudan con las operaciones y mantenimiento, sino que también dan a las organizaciones la oportunidad de valorar monetariamente sus datos y los de sus socios, así como los servicios desarrollados para monitorear diferentes componentes del sistema. El aprovechar esta nueva economía de datos y servicios amplía el alcance de los GD desde mejorar el mantenimiento y operación de un solo componente ciberfísico, por ejemplo, un motor de avión, hasta un enfoque más amplio, por ejemplo, un plano completo junto con la optimización de la operación de toda la flota. Con un alcance tan ampliado, la necesidad de desarrollar gemelos de gemelos es cada vez más frecuente.

Todos estos escenarios están impulsados por el objetivo de recopilar tantos datos relevantes como sea posible para conducir una toma de decisiones oportuna. Aprovechar los gemelos digitales requiere el diseñarlos teniendo en cuenta estos diversos objetivos de alcance. Los datos, la arquitectura, el diseño de la infraestructura de recopilación y servicio de datos,

y el modelo digital del activo son todos parte del alcance.

Además de las partes interesadas típicas, como usuarios, equipos de desarrollo y clientes, con GD las partes interesadas de los sistemas incluyen científicos de datos y analistas del dominio del negocio, quienes crean sus propios pasos de análisis mediante el uso de plataformas y aplicaciones de interfaces de programación para GD, acceso a los datos y otras capacidades analíticas. Algunos también pueden asumir el rol especial de constructor de gemelos – aquellos que crean gemelos digitales a partir de modelos y análisis digitales existentes.

Los intereses (concerns) arquitectónicamente significativos se pueden agrupar en tres áreas: datos, funcionamiento y evolución-mantenimiento. Esta clasificación se aplica en cualquier caso, ya sea el gemelo digital un elemento físico, componente de software, proceso de negocio o servicio. Obtener y expresar estos intereses como escenarios de atributos de calidad. Ayudaría a los equipos de ingeniería de software a evaluar tecnologías y sus riesgos a la hora de diseñar sistemas que incorporen gemelos digitales.

Intereses de calidad basados en datos para diseñar GD incluyen escalabilidad, consistencia de un activo o ciclo de vida del proceso, la accesibilidad, la usabilidad de la analítica de los datos, seguridad y privacidad. Hay otros atributos críticos relacionados con los datos, como la propiedad y propiedad intelectual, el control de versiones y ciclos de vida (por ejemplo, cuánto tiempo resguardar la información). No todas estas características pueden ser controladas con estrategias arquitectónicas; sin embargo, conocerlas informa a los arquitectos sobre otras limitaciones potenciales a considerar al seleccionar tecnologías y enfoques de diseño.

Los *intereses funcionales u operativos* están relacionados con cómo funciona el gemelo digital así como cuándo y cómo se realiza el análisis

de datos. Aquí son relevantes los problemas de rendimiento, disponibilidad, seguridad y escalabilidad, especialmente cómo se recopilan los datos operativos mientras el sistema funciona. Diferentes medidas de la calidad pueden aplicarse al gemelo y al activo en operación. La recopilación y servicio de datos en tiempo real, que pueden ser utilizados para orquestar escenarios como reparación en tiempo real y disponibilidad, solicitan estrategias de diseño más estrictas, comparando con sistemas tradicionales.

Los intereses relacionados con *la evolución y el mantenimiento* se basan en considerar la evolución del activo que el gemelo digital representa así como la evolución de las tecnologías en las que se basa su arquitectura. La integrabilidad heredada, la reutilización, la facilidad de configuración y la alineación de la plataforma con el gemelo digital, así como las consideraciones relacionadas con su evolución impulsan diferentes clases de decisiones arquitectónicas. En particular, la capacidad de evolución es a menudo una cualidad de mayor prioridad y preocupación en los sistemas habilitados para GD, debido al rápido ritmo de cambio en las tecnologías subyacentes así como en los datos.

Las decisiones de diseño de software para GD operan en dos niveles relacionados pero distintos. El primero es la arquitectura del gemelo digital, en esencia, las elecciones que están involucradas en cómo modelar el activo. El segundo nivel se refiere a las decisiones relacionadas con la incorporación del gemelo digital en el sistema, particularmente considerando que la operación y las opciones de mantenimiento y evolución se realizan en función de los datos recopilados.

Si te interesa el tema, puedes empezar viendo la edición especial de la revista IEEE Software de Marzo/Abril 2022⁴, con varios artículos dedicados, incluso el artículo completo del que se ha extraído este resumen.

³<https://link.springer.com/article/10.1007/s10816-021-09518-6>

⁴<https://www.computer.org/csdl/magazine/so/2022/02>



