

Noticias del Departamento de Ingeniería de Sistemas

Novedades



Presentación de ponencias en Workshop de Investigadores en Ciencias de la Computación 2021:

- **Reusabilidad en el contexto de Desarrollo de Sistemas para Big Data**
- **Clasificando información en sitios de CQA**

¿Qué es el WICC?

El Workshop de Investigadores en Ciencias de la Computación (WICC) es organizado, a partir de 1999, por la Red de Universidades Nacionales con Carreras de Informática (RedUNCI) con el propósito de generar un foro para el intercambio de ideas entre investigadores en Ciencias de la Computación, de modo de fomentar la vinculación y potenciar el desarrollo coordinado de actividades de Investigación, Desarrollo e Innovación entre los mismos.

El objetivo central es reunir líderes de proyectos, investigadores y becarios de I+D+i, de todas las universidades miembro de la RedUNCI, de modo de discutir las líneas de investigación y los posibles mecanismos y temas/áreas específicas de cooperación. WICC es un espacio abierto a la comunidad (docentes, investigadores, alumnos) de modo de favorecer la difusión de los temas y

las relaciones entre los actores de las actividades de IDi en las Universidades.

Las ponencias presentadas fueron las siguientes:

Reusabilidad en el contexto de Desarrollo de Sistemas para Big Data

por AGUSTINA BUCCELLA, JUAN LUZURIAGA, ALEJANDRA CECHICH, LÍAM OSYCKA, FACUNDO PATERNO, MATÍAS POL'LA, MARCOS CRUZ, RODOLFO MARTÍNEZ, RAFAELA MAZALU, MARCELO MOYANO



A pesar de que existen mecanismos para manejar el cambio de esquemas y datos en un modelo relacional, alcanzar un nivel similar en Big Data es todavía un desafío importante debido a la naturaleza flexible de los esquemas de almacenes NoSQL.

En trabajos relacionados, se presentan seis arquitecturas de referencia¹ propuestas actualmente en la literatura para SBDs y se analiza el cumplimiento de las cinco Vs como requerimientos tradicionales. El análisis se hace de manera transversal; es decir, se analiza cada requerimiento por separado y luego sus interacciones con otros requerimientos. Por ejemplo, en el contexto de SBDs se explotan almacenes NoSQL que guardan datos con bajo acoplamiento y flexibles; sin embargo, no satisfacen requerimientos de Variabilidad y Veracidad, por lo que, en la práctica, las aplicaciones resultan en modelos ad hoc. De ahí la necesidad de contar con arquitecturas de referencia

¹Una arquitectura de referencia combina el conocimiento sobre arquitecturas en general con la experiencia en requerimientos específicos de una solución en un dominio de problema.



que faciliten el desarrollo concreto, conociendo los componentes y sus relaciones previamente para permitir el análisis de propiedades. Del análisis realizado, se desprende que sólo una propuesta satisface los requerimientos de Variabilidad por medio de (1) el almacenamiento del esquema de información de los elementos incorporados; (2) la existencia de estadísticas descriptivas para acceder a la evolución de los datos; y (3) el almacenamiento de la información sobre fuentes de datos usando un repositorio de metadatos (MetaData Management system).

Por otra parte, existen antecedentes donde se presenta el desarrollo de una arquitectura para un SBD en un caso específico, haciendo uso de una arquitectura de referencia que define una familia de sistemas relacionados. En particular, como el dominio es demasiado amplio, la arquitectura de referencia se ve acotada por medio de casos de uso (ej. visualización y análisis de información geoespacial estratégica, análisis inteligente de señales, etc.). Esta arquitectura de referencia sirve como mecanismo para capturar y compartir conocimiento, conteniendo tanto conocimiento del dominio (casos de uso) como conocimiento de la solución (la correspondencia a tecnologías concretas).

Considerando estas propuestas, nuestro trabajo se enfoca en incorporar Variabilidad en SBDs a través del modelado de elementos reusables de un dominio – sea este de negocios o tecnológico. Esto nos lleva a incorporar información (y sus posibles usos), a modo de línea de productos software (LPS).

En particular, los desarrollos en LPS se centran en identificar similitudes y variabilidades dentro de dominios particulares para ser reutilizados cuando se construyen nuevos productos. A su vez, este reuso de dominios puede ser extendido a subdominios, en especial cuando entre ellos existen relaciones o aspectos similares.

Clasificando información en sitios de CQA

por VALERIA ZORATTO, GABRIELA ARANDA, NADINA MARTÍNEZ CAROD, CARINA NODA, MAURO SAGRIPANTI



La cantidad de información disponible en Web crece día a día. En particular, los servicios de Community Question Answering (CQA) se han convertido en una forma popular de búsqueda de información en línea, donde los usuarios pueden interactuar e intercambiar conocimientos en forma de preguntas y respuestas. La información que contienen los foros de discusión y las CQA son muy valiosas para usuarios con los mismos intereses y más aún, cuando tienen un problema similar, ya que estos sitios son colaborativos, donde se presentan discusiones e intercambio de ideas sobre un tema específico.

Una de las características que distingue a los foros de discusión es que su interacción es asincrónica, es decir, que no se necesita estar conectados al mismo tiempo para obtener la información que solucione un problema. Incluso en la gran mayoría de las veces los participantes no se conocen personalmente, pero sí a través de sus nombres, alias o avatares (representaciones gráficas que se asocian a usuarios para identificarlos). Un ejemplo de estos foros puede ser Yahoo Answers! (YA). YA es un sitio web de preguntas y respuestas impulsado por la comunidad o un mercado de conocimiento de Yahoo!, que permite a sus usuarios tanto formular preguntas como responderlas. Otro foro a destacar es *Stackoverflow*, el cual está referido específicamente

a problemas en entornos informáticos. Ofrece además la posibilidad de agregar código como respuesta a una pregunta, e incluye información de los usuarios, por ejemplo, la reputación, que, cuantos más puntos tenga el usuario, más cosas puede hacer en la comunidad.

Un foro de discusión está compuesto por múltiples hilos. Los hilos son creados por usuarios que tienen algún problema o duda respecto a un tema. El usuario abre un hilo con una consulta inicial y a partir de ese momento, los usuarios de la comunidad podrán debatir en función del tema y del problema enunciado. En base a este comportamiento es que luego, para obtener conocimiento proveniente de los hilos de discusión se utilizan diferentes técnicas y estrategias para establecer cuáles de las posibles soluciones obtenidas de los foros pueden ser relevantes para consultas sobre problemas similares.

El proyecto realiza, por un lado, el tratamiento del texto contenido en los hilos dentro del foro y por otro lado analiza la red de usuarios que existe en él. También se ha estudiado la satisfacción del usuario y se intenta predecir si el autor de la pregunta estará satisfecho con las respuestas enviadas por los participantes de la comunidad.

En base a estos antecedentes, nuestro proyecto tiene como objetivo principal hacer reuso de la información existente en foros de discusión de la Web, haciendo uso no solo de la información textual de los hilos sino teniendo en cuenta además a los usuarios, utilizando la red que se forma con las distintas interacciones que tienen con el resto de la comunidad y analizando los roles que cumplen, para poder detectar usuarios expertos y darle un mayor peso a las respuestas candidatas de una pregunta.



Competitividad Digital



¿Usar recursos en la nube para evitar la extinción de especies?

La Universidad de Sydney (USyd)² espera salvar especies en peligro de extinción en Australia aprovechando el poder de cálculo de los servicios Web de Amazon³ (AWS) para acelerar la investigación en genomas, evitando así por ejemplo, la extinción del demonio de Tasmania. La misión del *Australasian Wildlife Genomics Group* en la USyd es ayudar a proteger especies en peligro de extinción, generalmente únicas en Australia, que se separó de otros continentes hace 95 millones de años y donde 87% de sus mamíferos, 93% de sus reptiles, 94% de sus ranas y 45% de sus aves son propios. Asegurar su supervivencia, es vital para la biodiversidad.

El demonio de Tasmania es un raro marsupial, que se encuentra sólo en la isla de Tasmania. Usando su genoma, los investigadores puede proveer a agencias de conservación con información científica detallada sobre cómo proteger mejor a esa especie. Sin embargo, trabajar con un genoma es como tener un rompecabezas de un billón de piezas, *sin tener ninguna guía para armarlo!* Eso implica un trabajo laborioso y largo, que ahora se acelera gracias a los recursos en la nube.

El trabajo en USyd se ha acelerado desde que comenzaron una prue-

ba con AWS Cloud, lo que permitió no sólo aumentar la velocidad de las investigaciones, sino el manejo más efectivo de los fondos. El poder computacional de AWS se acopló con la interface RONIN⁴ – una aplicación Web que permite a cualquiera utilizar recursos computacionales complejos – permitiendo procesar, analizar y categorizar todos esos datos, construyendo la imagen final.

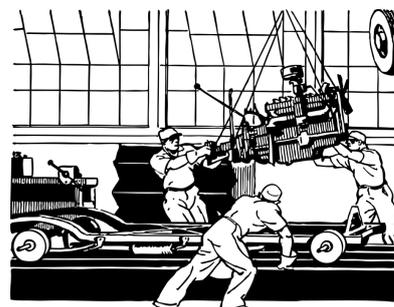
USyd compartirá los datos del genoma en el repositorio público de AWS (*AWS Public Dataset Program*), que es una iniciativa para que investigadores accedan a demanda a los depósitos de datos para acelerar descubrimientos científicos. Aún con información fragmentada, se pueden establecer patrones. Y generalmente, esos patrones se cumplen en otras especies, por lo que pueden usarse para sacar ciertas conclusiones o reglas.

El equipo de la USyd continua ensamblando y anotando los genomas de más especies. Los recientes incendios en Australia han sido catastróficos para la vida silvestre, empujando a más especies al peligro de extinción. En términos de escalar la propuesta en beneficio de otras especies, el objetivo final es crear una biblioteca de genomas universal y herramientas que otros investigadores y conservacionistas puedan acceder para tomar decisiones de carácter científico.

Si te interesa el tema de aplicaciones posibles en conservación de especies, hay miles de ellas!!! Por ejemplo, se puede ver cómo la Inteligencia Artificial y el uso de drones intentan salvar la Gran Barrera de Coral⁵; cómo IoT está revolucionando la conservación de la vida silvestre⁶ o salvando especies en peligro⁷, como elefantes en África⁸.

Mesa del Arquitecto

¿Para qué sirve una Línea de Productos Software?



La *Ingeniería de Líneas de Productos de Software* (ILPS) contempla la definición de partes comunes de un conjunto de productos dentro de un dominio específico y a la vez provee un mecanismo para modelar la variabilidad dentro de los mismos. En general esta ingeniería persigue beneficios como la reducción de los costos de desarrollo, aumento de la calidad, reducción del tiempo de puesta al mercado (Time to Market), reducción del esfuerzo de mantenimiento, mejora en la estimación de costos, etc.

A diferencia del desarrollo de software individual, en la ILPS se debe realizar un mayor esfuerzo inicial para identificar y obtener aquellos requerimientos que formarán parte de la línea, es decir, la plataforma de los productos. Estos requerimientos son tanto funcionales como no funcionales, incluyendo particularmente los atributos de calidad de un sistema de software. Una vez obtenida la plataforma, se deben identificar los requerimientos particulares de cada producto. Es en este punto, donde la *flexibilidad de la plataforma* juega un rol fundamental ya que es la que definirá, en cierta medida, el rango de productos que es posible crear desde la *Línea de Productos de Software* (LPS).

²<https://dpiwpe.tas.gov.au/wildlife-management/save-the-tasmanian-devil-program>

³<https://aws.amazon.com/es/>

⁴<https://ronin.cloud/>

⁵<https://www.techrepublic.com/article/how-ai-and-drones-are-trying-to-save-the-great-barrier-reef/>

⁶<https://www.perle.com/articles/how-the-iot-is-revolutionizing-wildlife-conservation-40189360.shtml>

⁷<https://www.nec.com/en/global/insights/article/2020022502/index.html>

⁸<https://www.techrepublic.com/article/how-iot-tech-is-helping-african-rangers-protect-endangered-elephants-from-poachers/>



Dependiendo de la metodología de LPS seguida, se aplican diferentes mecanismos de acuerdo con las dos fases que intervienen en todo desarrollo de ILPS: la ingeniería de dominio y la ingeniería de aplicación.

En la primera fase, el análisis de dominio está dirigido a identificar, capturar y organizar toda la información fuente recopilada de los sistemas existentes, los expertos del dominio, libros de texto, etc. Como resultado, se obtiene una plataforma de software con un conjunto de componentes reusables y configurables que proveen un entorno en común. A su vez, dicha plataforma posee un conjunto de variabilidades que permiten justamente adaptarla a las necesidades particulares de las aplicaciones a ser creadas.

En la segunda fase, la ingeniería de aplicación, se instancia dicha plataforma, mediante la configuración de la variabilidad definida. Finalmente, la última actividad de esta fase devuelve un producto de software en particular.

Es deseable que la base que contiene las características comunes sea lo más flexible posible para permitir el reuso de la misma, y lo suficientemente acotada para no generar un conjunto inmanejable de posibilidades. De esta manera, la clave principal de todo desarrollo de LPS, es que las técnicas y metodologías que se proponen promuevan la flexibilidad (por medio de la variabilidad). Esta flexibilidad, asociada a la plataforma, debe permitir que los módulos o componentes que la componen sean adaptables a los diferentes sistemas que puedan ser generados. Para lo-

grar esto, es necesario identificar y describir las funcionalidades de la línea de productos que pueden variar en función de las características que ofrecen, los requisitos que necesitan, o incluso en términos de la arquitectura subyacente. Es por esto que la plataforma de la LPS deberá proporcionar los medios para satisfacer las necesidades de diferentes clientes.

Una forma de propiciar un ambiente para mejorar la flexibilidad de la LPS es por medio de la definición y uso de componentes bien definidos en base a los lineamientos de la *Ingeniería de Software Basada en Componentes* (ISBC). También es necesaria la adopción o redefinición de una metodología para la gestión de variabilidad que permita administrar la variabilidad durante todo su ciclo de vida.

Como se puede observar a partir de estas dos fases, el éxito del desarrollo de una LPS depende de la identificación, uso y administración de los artefactos reusables. Por lo tanto, la aplicación de técnicas específicas para la sistematización del reuso es crucial aquí.

Si te interesa el tema, que es realmente extenso, existe numeroso material bibliográfico en libros y estándares. Por ejemplo, el libro de K. Pohl, G. Böckle, F. Van Der Linden, *Software Product Line Engineering: Foundations Principles and Techniques*, Springer, 2005; el modelo de referencia propuesto en el estándar ISO/IEC 26550:2015: “Software and systems engineering — Reference model for product line engineering and management”⁹; material disponible en el repositorio del Software Engineering Institute (SEI)¹⁰; o la edición especial de la revista *IEEE Software* de Ma-

yo/Junio 2010 “Successful Software Product Line Practices”, donde también se discuten experiencias al combinar orientación a servicios con líneas de productos software.

¿Sabías qué ...

... se puede tener una certificación como instructor CMMI?

La organización ISACA¹¹ es una asociación profesional internacional enfocada en el gobierno de T&I (tecnología e información). ISACA a nivel mundial engloba dos ramas – el ITGI (Information Technology Governance Institute) y el CMMI Institute.



Los cursos de entrenamiento que ofrece ISACA están enmarcados en programas de certificación, entre ellos y referidos a CMMI: *CMMI Associate*, *CMMI Professional* y *CMMI Partner*¹². Después de elegir el programa de certificación, se seleccionan cursos de entrenamiento que sean afines a la categoría (*build, advance, master*); por ejemplo, *CMMI Instructor Training - build*, *Advancing Organizational Capability: Applying CMMI - Advance*, *Lead Appraiser Training - Master*¹³.

Si te interesa, el programa cuenta con descripción más detallada en el sitio Web, incluyendo precios y modalidad de cada curso.

⁹<https://www.iso.org/standard/69529.html>

¹⁰<https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=423718>

¹¹<https://cmmiinstitute.com/>

¹²<https://cmmiinstitute.com/learning/certifications>

¹³<https://cmmiinstitute.com/learning/training>

