

Síntesis Tesis

TESIS DE MAGÍSTER EN INGENIERIA DE SOFTWARE

“Metodología para el descubrimiento de conocimiento en bases de datos: un estudio comparativo”

Juan Miguel Moine

Directora: Dra. Ana Silvia Haedo
CoDirectora: Dra. Silvia Gordillo.

Motivación

Para llevar a cabo en forma sistemática el proceso de descubrimiento de conocimiento en bases de datos, conocido como minería de datos, es necesaria la implementación de una metodología. Actualmente las metodologías para minería de datos se encuentran en etapas tempranas de madurez, aunque algunas como CRISP-DM ya están siendo utilizadas exitosamente por los equipos de trabajo para la gestión de sus proyectos. En este trabajo se establece un análisis comparativo entre las metodologías de minería de datos más difundidas en la actualidad. Para lograr dicha tarea, y como aporte de esta tesis, se ha propuesto un marco comparativo que explicita las características que se deberían tener en cuenta al momento de efectuar esta confrontación.

El objetivo principal de este trabajo de tesis es la construcción de un marco comparativo que sirva como herramienta para poder evaluar y confrontar diferentes metodologías de minería de datos. El objetivo secundario de esta investigación es el análisis mediante un estudio descriptivo-comparativo de las metodologías más difundidas en la actualidad para proyectos de minería de datos, abriendo una discusión sobre qué enfoques deberían ser realmente considerados una metodología.

Aportes de la tesis

En este trabajo de tesis se ha logrado la construcción de un marco comparativo como herramienta para la confrontación de metodologías de minería de datos. El marco propuesto incluye cuatro aspectos donde se analiza el nivel de especificación de las tareas, los escenarios de aplicación, las actividades que componen cada fase del proceso y las actividades destinadas a la dirección del proyecto. Para cada uno de estos aspectos se propone la evaluación de un conjunto de características que deberían estar presentes en una metodología de minería de datos bien definida. El marco comparativo ha sido utilizado para confrontar las metodologías CRISP-DM y Catalyst. Aunque ambos enfoques se encuentran actualmente en etapas tempranas de madurez, han logrado cumplir un gran porcentaje de las características evaluadas. Sin embargo, durante el estudio también se han evidenciado los puntos que se deberían mejorar y seguir desarrollando, como las actividades destinadas a la dirección del proyecto. Si bien en este trabajo se han analizado los cuatro aspectos en función de la proporción de características que se cumplen en cada uno, los resultados podrían ser diferentes si se trabajara con puntajes. Si el usuario del marco comparativo lo considera necesario, puede efectuar dicha valoración según su criterio. Se espera que el resultado de esta tesis sirva como herramienta para que los equipos de trabajo puedan evaluar metodologías de minería de datos, considerando además la posibilidad de complementar las mismas con los conceptos que no estén presentes o bien utilizar el marco como base para construir una propia.

Líneas de I/D futuras

Como línea de trabajo futuro se podría establecer una ponderación de las características evaluadas mediante un sistema de puntajes que permita obtener resultados cuantitativos respecto al desempeño de cada metodología. El desafío de esta investigación radicará en los criterios que se deberían tener en cuenta para establecer estos valores en forma objetiva. Otra línea de investigación podría ser la ampliación del marco comparativo, incorporando nuevos aspectos al mismo.

Síntesis Tesis

TESIS DE DOCTORADO EN CIENCIAS INFORMATICAS

“Un entorno de aprendizaje y una propuesta de enseñanza de Simulación de Eventos Discretos con GPSS”

Gonzalo Luján Villarreal

Directora: Ing. Marisa R. De Giusti

CoDirectora: Dra. Silvia Gordillo.

Motivación

La enseñanza en el área de simulación de eventos discretos requiere integrar una variedad de conceptos teóricos y ponerlos en práctica a través de la creación y ejecución de modelos abstractos de simulación, con el objetivo de recopilar información que pueda traspolarse hacia los sistemas reales. Para construir modelos, ejecutarlos y analizar los resultados de cada ejecución se utilizan herramientas de software cada vez más sofisticadas que permiten expresar los elementos de los modelos en términos de entidades abstractas y relaciones, y que recopilan gran cantidad de datos y estadísticas sobre cada una de estas entidades del modelo. GPSS es una de estas herramientas, y se compone de un lenguaje de programación por bloques y un motor de simulación que traduce estos bloques en distintas entidades del modelo. A pesar de que su primera versión data de 1961, GPSS es aún muy utilizado por profesionales y empresas, y es una de las herramientas más utilizadas para la enseñanza de simulación de eventos discretos por instituciones académicas de todo el mundo. El avance de la capacidad de cómputo de las computadoras ha permitido incorporar una mayor cantidad de herramientas y funciones a las distintas implementaciones de GPSS. Mientras que esto representa una ventaja para sus usuarios, requiere también un cada vez mayor esfuerzo por parte de los docentes para enseñar a sus estudiantes a aprovechar todo su potencial. Muchos docentes e investigadores han buscado optimizar la enseñanza de simulación de eventos discretos desde múltiples ángulos: la organización del curso y la metodología de enseñanza, la creación de elementos de aprendizaje que ayuden a aplicar los distintos elementos teóricos, la generación de herramientas para construir modelos GPSS, y la construcción de herramientas para comprender el motor de simulación por dentro.

Aportes de la tesis

En esta tesis se introduce una herramienta de software que permite construir modelos GPSS de manera interactiva, cuyo diseño fue pensado para integrar los elementos teóricos del curso con los objetos y entidades de GPSS. Esta herramienta también permite ejecutar estos modelos y analizar con alto nivel de detalle su evolución a través del tiempo de simulación, lo que permite a los estudiantes comprender cómo funciona el motor de simulación y cómo interactúan las distintas entidades entre sí. Se incluye también una propuesta de enseñanza basada en una fuerte participación de los estudiantes, que, por medio de esta nueva herramienta, les permite incorporar los conceptos más fácilmente. Esta propuesta de enseñanza fue puesta a prueba con alumnos del área de sistemas, quienes tomaron un curso que contiene los mismos elementos teóricos y prácticos de un curso tradicional, pero con una organización diferente. Entre los resultados logrados se destacan una reducción del tiempo requerido para aprender los conceptos de GPSS cercana al 50%, una mayor capacidad por parte de los alumnos para asimilar conceptos y derivar nuevos conceptos por sí solos, a partir de conceptos adquiridos previamente.

Líneas de I/D futuras

Una de las líneas de trabajo futuras implica la ampliación de las herramientas de reporte, haciendo hincapié en el historial de simulación de cada entidad, y en la navegación entre entidades y a través de los tiempos de simulación. Desde el punto de vista del curso, deberá analizarse la aplicabilidad del método en un curso de mayor tamaño por un lado, y en un curso a distancia -con sus respectivas adaptaciones- por el otro. Las posibilidades de uso de esta herramienta de manera virtual y la correspondiente adaptación del método propuesto permitirán estudiar el aporte de este trabajo sobre un conjunto mucho más amplio y heterogéneo de individuos. Por otro lado, resulta también muy interesante la integración de esta herramienta con entornos de aprendizaje (LMS), ya que permitirá al docente preparar los materiales del curso de enseñanza y ofrecerlos en el contexto del espacio de trabajo. Asimismo, la incorporación de una interfaz de creación de modelos predefinidos para el docente ofrecerá la posibilidad de generar actividades de análisis, inspección e investigación, que podrán vincularse con los materiales de estudio así como también con los conceptos teóricos vistos en clase.