UNIVERSIDAD NACIONAL DE SAN LUIS

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS Y NATURALES

Tesis
para optar a la titulación de postgrado correspondiente a la
Maestría en Ingeniería de Software

Una Metodología para el Desarrollo de Aplicaciones Autoadaptativas basada en Business Intelligence. Aplicación en Medicina

Marcelo Fabio Roldán

Directores de Tesis: Mg. Ing. Germán Montejano
Mg. Ing. Ana Funes

San Luis
2012

Prólogo

A partir de las obras de Fayaad, Shapiro-Pietrostsky, Smith y Uthurusamy [BRA98] [FAY96], quienes en 1996 acuñaron el término de KDD como búsqueda de conocimiento en grandes bases de datos, el mundo ha experimentado cambios en lo que respecta a la manera de explorar el conocimiento oculto.

El devenir de los años, con el incremento constante de la tecnología, ha provocado aumentos exponenciales en el volumen de la información, que hacen imposible para el humano realizar esta búsqueda con métodos tradicionales.

Para dar solución a este nuevo paradigma, el uso de las técnicas de aprendizaje de máquina, inteligencia artificial, estadística avanzada, computación neuronal y otras disciplinas emergentes, se han vuelto de uso cotidiano.

Análogamente a estos avances, el almacenamiento de datos comienza a producir modelos de uso que permiten no solo el almacenamiento de enormes volúmenes de información - en el orden de los petabytes – sino también facilitan su acceso para el análisis a través de Internet, brindando de este modo una expansión impensada unos pocos años atrás.

Todos estos cambios han requerido de una tecnología integradora que permita la utilización colaborativa de estas herramientas y que brinde resultados que permitan optimizar procesos, mejorar eficiencias, aumentar productividad, mejorar finanzas, evitar fraudes, entre muchas otras aplicaciones. Estos son los orígenes del Business Intelligence.

Sin embargo, en este entorno del Business Intelligence, quedan desprotegidos aquellos proyectos, en los cuales el factor económico no representa de modo sustancial el núcleo de la existencia de una línea de investigación.

Amén de todo esto, el aprendizaje logrado a partir de los datos no constituye el cenit de los objetivos, ya que los datos, tales como sucede en el mundo real, cambian, se actualizan, migran entre fronteras y también desaparecen, entonces, ¿Cómo seguirles la huella en este marco cambiante?

La adaptabilidad como fuente de evolución aparece como la mágica tecnología que emula la inteligencia al proveer un artilugio, el que además de considerar los patrones que se hallan en los bancos de datos, será capaz luego de un tiempo, de reacomodarse, aprender nuevamente y optimizarse para su uso.

Ante este escenario, quedan pendientes las definiciones de metodologías, que faciliten la producción de conocimiento a partir del aprendizaje y del descubrimiento, en especial cuando se trata de aquellas áreas de las ciencias, como las médicas, farmacéuticas, génicas o más inclusivamente llamadas ciencias de la vida, que conceptualmente se pueden considerar como alejadas del espíritu del Business Intelligence o del Adaptive Business Intelligence.

Este trabajo plantea una solución eficaz a este dilema, a través de un itinerario donde se aborda el estado actual, sus concepciones teóricas, una serie de aplicaciones resueltas y un contraste con las

metodologías existentes de mayor uso. La resultante es una metodología moderna, actualizada y adecuada, para su uso en áreas específicas de las ciencias con una fuerte visión hacia el Adaptive Business Intelligence.

Ya desde su gran obra, Michaelewicsz [MIC07], parangonando el proceder de las aplicaciones adaptativas con las del magistral detective Sherlock Holmes, nos enseña que estas aplicaciones han venido para quedarse.

Resumen

La búsqueda constante por encontrar causas de enfermedades y mejorar las expectativas de vida de sujetos enfermos, así como la creciente innovación en especial en materia del desarrollo de nuevas tecnologías orientadas a la inteligencia artificial, a la minería de datos y a la extracción de conocimiento, sitúan al experto en sistemas en un entorno hostil debido a la variedad de herramientas con las que cuenta. Más aún se hace manifiesto en razón de la escasez de conceptos metodológicos claros en relación a su aplicación directa en campos como la medicina.

Por consiguiente, el problema abordado facilita la identificación práctica de patrones característicos y distintivos de la base de datos médica, lo que permite una diferenciación clara de las etapas metodológicas a seguir para realizar aplicaciones adaptativas en el ámbito de la medicina.

Para esto se realiza una aproximación desde el enfoque amplio que proporcionan algunas metodologías existentes orientadas al Business Intelligence, y se toman como referencia diversos casos prácticos, se clasifican y ordenan las etapas más significativas que estos procedimientos deben seguir para obtener un producto aplicable a esta rama específica de las aplicaciones de los sistemas de información.

La información que brinda este estudio, se espera sirva para facilitar la realización ágil de aplicaciones adaptativas a través de esta metodología simplificada.

Asimismo, para validar estas premisas, se emplea esta propuesta metodológica mediante el desarrollo de dos aplicaciones autoadaptativas orientadas al profesional médico, que le sirve para la toma de decisiones en dos casos específicos tomados como referencia.

Se consideran aspectos de metodologías aplicadas actualmente, enfocados por sus autores desde puntos de vista dispares. Por esto se hace menester adecuarlas a la necesidad particular de las aplicaciones en la medicina así como en las ciencias de la vida. En estas disciplinas prevalecen las excepciones, evaluaciones estadísticas previas de los datos, certeza relativa en los resultados de predicción entre otras características propias.

La generalización de las conclusiones obtenidas a partir de estos desarrollos de Business Intelligence, permiten deducir su aplicación real a otros campos y disciplinas médicas, con enfermedades tan diversas como diabetes, dermatitis, problemas de nutrición, epidemiología, aplicación de drogas entre otras.

Agradecimientos

"El que mucho abarca, poco aprieta", este viejo refrán que mi padre solía repetirme de niño ha tomado vigencia en este trabajo, siendo el espíritu para abordar el análisis crítico de las metodologías que rigen actualmente la tecnología de Data Mining, en sus formas de SEMMA, KDD Process o CRISP-DM.

De esta manera un refrán, que aunque no es compartido en lo personal, ha sido la fuente inspiradora para encontrar el talón de Aquiles de estas construcciones de la industria, ya que para problemas específicos no son eficientes por su generalidad.

Esta generalidad se transmuta en un conocimiento disciplinar cuyos beneficios se espera, logren difundirse de manera efectiva, ya que el auge de esta tecnología, teniendo en cuenta las ciencias hacia las cuales está dirigida, seguramente representarán un beneficio a la humanidad.

Y en especial …

A mi esposa

Y a mis hijas,

Por las horas de felicidad robadas para concluir este trabajo.