

Modelo para la estimación temprana de esfuerzo en proyectos de software, incorporando información de proyectos similares

Ana Lucía Pérez, Lilibiana González, Astrid Duque, Felipe Millane
y Germán Ospina

UNIVERSIDAD DE ANTIOQUIA.
Departamento de Ingeniería de Sistemas
alperez@ulea.edu.co

Recibido para revisión May-2006, aceptado Jun-2006, versión final recibida Jun-2006

Resumen: Existen modelos para estimación temprana del esfuerzo requerido en el desarrollo de proyectos software que retornan valores muy precisos si se utilizan en fases avanzadas del desarrollo, pero que presentan limitaciones si se usan en las fases iniciales. En este artículo se presenta un modelo que permite estimar durante las primeras fases del desarrollo, la capacidad de los analistas de la Gerencia Informática de ORBITEL para la creación de soluciones y la operación del negocio. De esta manera se busca facilitar la estructuración de planes de trabajo en los que se defina la disponibilidad del recurso humano para el desarrollo de soluciones informáticas. El modelo expuesto en este artículo es parte de los resultados obtenidos en un proyecto de investigación aplicada realizado entre el Departamento de Ingeniería de Sistemas de la Universidad de Antioquia y Orbitel S.A. con el objetivo de modelar la capacidad que tiene la Gerencia Informática para crear y operar soluciones basadas en tecnologías de información. Considerando los históricos disponibles en Orbitel, nuestro modelo entrega estimaciones precisas desde el punto de vista estadístico. Adicionalmente, el modelo propuesto es simulado con una herramienta orientada a la Web.

Palabras Clave: Estimación temprana, modelo de estimación de esfuerzo en proyectos de software.

1 INTRODUCCIÓN

Dentro de la gestión de un proyecto de desarrollo de software, la planificación es una actividad de gran importancia, en la que se establecen objetivos y metas de un proyecto, además de las estrategias, políticas y procedimientos para alcanzarlos [Barbacci (1985)].

Pero, los proyectos que involucran Tecnologías de Información (TI) tienen un alto grado de incertidumbre, su comportamiento puede variar con el tiempo y de acuerdo a variables como el recurso humano y la naturaleza de la empresa en la que se desarrollen, y además, las decisiones tomadas en un instante determinado por quienes se encuentran en el nivel estratégico de la organización, afectan la evolución del proyecto de diversas maneras, directas o indirectas [Toro (2002)]. Estas características complejas dificultan la labor de planificación inicial, realizada por los gestores de proyectos.

Una de las variables a calcular durante la planificación del proyecto es el esfuerzo, es decir, la fuerza de trabajo requerida para el desarrollo, medida en meses-hombre, días-hombre, y en general, unidad de tiempo-hombre [Zhenyou (2004)]. Este cálculo debe hacerse lo

más temprano posible, idealmente, cuando se tiene la especificación de requisitos, porque antes de esto cualquier estimación será solo una vaga aproximación a la realidad.

En la Gerencia Informática de Orbitel es necesario estimar el esfuerzo para la gestión de un proyecto antes de tener la especificación de requisitos completa. El proyecto de investigación aplicada "MODELO PARA IDENTIFICAR LA CAPACIDAD DE CREAR Y OPERAR SOLUCIONES BASADAS EN TECNOLOGÍAS DE INFORMACIÓN" apunta a la creación de un modelo para estimación temprana de esfuerzo a partir de las variables asociadas al proceso de desarrollo.

La estructura del artículo se indica a continuación: En la Sección 2 se presentan las técnicas utilizadas en la actualidad para resolver el problema de calcular el esfuerzo de acuerdo al tamaño del producto software a construir; en la Sección 3 se describen las fases ejecutadas para obtener el "Modelo de capacidad", un nuevo método de estimación temprana de esfuerzo; posteriormente, en la Sección 4 se muestran los resultados obtenidos con el "Modelo de capacidad"; y por último, en la Sección 5 se presentan algunas conclusiones.

2 TÉCNICAS PARA EL CÁLCULO DEL ESFUERZO DE ACUERDO AL TAMAÑO DEL PRODUCTO

El principal factor que influye en el cálculo del esfuerzo es el tamaño del producto a desarrollar, es por esto que se han propuesto métodos diferentes del juicio experto que buscan disminuir la incertidumbre en la estimación del tamaño [Jorgensen y Sjoberg (2001)]. Entre las técnicas más estructuradas se encuentran: puntos función, puntos característica, puntos de casos de uso, entre otras. Cada una de estas técnicas tiene fórmulas para calcular el esfuerzo de acuerdo al tamaño del producto a construir.

La técnica de Puntos Función o Function Points [Albrecht (1979), Albrecht y Gaffney (1983)] proporciona una unidad de medida para la funcionalidad de los sistemas software determinando sus componentes principales: entradas, salidas, consultas o peticiones interactivas (cuando el usuario hace una petición al sistema y éste devuelve una respuesta), archivos lógicos internos (archivos maestros) y archivos lógicos externos (interfaces con otras aplicaciones), y luego asociando estos componentes a características generales de un sistema (eficiencia, reusabilidad, facilidad de operación y mantenimiento entre otras) [Sánchez (1999)].

La técnica ya descrita fue pensada para medir el tamaño funcional de sistemas software orientados a la gestión, pero, era necesario contar con una técnica útil para medir el tamaño funcional de otras aplicaciones. Con esta intención, Jones (1999) desarrolló una técnica experimental, denominada Puntos Característica (Feature Points), para adaptar la técnica de puntos función a sistemas software científicos y de ingeniería.

Los puntos característica se han venido utilizando con gran éxito en la medición de diversos sistemas software: sistemas en tiempo real, sistemas embebidos, software para inteligencia artificial, los cuales se caracterizan por la complejidad algorítmica que implementan y el escaso número de entradas y salidas que tienen. La técnica de Puntos de Casos de Uso [Peralta (2004), Ribu (2005)] permite determinar el tamaño de una aplicación de acuerdo al número de actores y casos de uso involucrados. Luego permite refinar este cálculo teniendo en cuenta el factor de complejidad técnica y el factor de ambiente en el cual se incluyen factores como las habilidades y el entrenamiento del grupo involucrado en el desarrollo.

Las técnicas presentadas anteriormente tienen como ventaja la confiabilidad en los resultados si se tiene toda la información requerida, pero, no son aplicables a proyectos de TI en general, y sólo se limitan a proyectos de desarrollo software, además, necesitan gran cantidad de datos que aún no se tienen en fases tempranas del desarrollo y trabajan sobre la base de una especificación de

requisitos buena, es decir, no ambigua y bastante completa [Varas (2002)].

Luego de una rigurosa revisión de la literatura para el estudio de la problemática de la Gerencia Informática de Orbitel, se llegó a la conclusión de que las técnicas existentes no permiten estimar el esfuerzo de los proyectos a realizar antes de tener una completa especificación de requisitos, por lo tanto, se propone a continuación un nuevo método para estimar el esfuerzo en etapas tempranas.

3 UN NUEVO MÉTODO PARA CALCULAR EL ESFUERZO EN ETAPAS TEMPRANAS

Esta investigación tiene lugar en la gerencia informática de Orbitel S.A, la cual cuenta actualmente con una estructura conformada por dos direcciones y 11 equipos de trabajo. La Dirección de Arquitectura y Soluciones Informáticas está integrada por los equipos: Arquitectura e Integración (AEI), Business Intelligence (BI), Ventas y Servicio al Cliente (VSC), Sistemas Administrativos y Financieros (SAF), Producto, Facturación, Plataformas Afines a Red (PAR). La Dirección de Infraestructura Informática y Atención a Usuario Final está integrada por los equipos: Planeación y Consecución de Recursos (PCR), Prestación de Servicios (PS), Gestión de la Relación con el Cliente (GRC), el cual no está ligado a ninguna de las direcciones ya mencionadas.

Cada uno de los equipos sigue un mapa de procesos bien establecido y trabaja bajo un esquema de autogestión que se rige por una política de hechos y datos en la cual cada procedimiento debe estar soportado y documentado. Por esto, el modelo para la estimación temprana de esfuerzo en proyectos de desarrollo de software, cobra gran importancia en la definición de indicadores y métricas que facilitan la planeación y control de las actividades que se realizan.

Esta investigación toma como punto de partida la información de proyectos terminados, de los cuales existen datos históricos de esfuerzo. En conjunto con los analistas y directores de la gerencia informática y haciendo uso tanto de la experiencia, como de técnicas estadísticas se logra tipificar y caracterizar los proyectos, con el fin de lograr primero, un modelo capaz de sugerir intervalos de esfuerzo en etapas tempranas para un nuevo proyecto y segundo, una herramienta de simulación que permita predecir, dado el esfuerzo estimado para cada nuevo proyecto, el comportamiento de la capacidad y los niveles de sobreesfuerzo de los analistas, las direcciones y en general de la gerencia informática.

Para la creación de este modelo se desarrollaron las siguientes actividades: estudio y tratamiento de datos históricos, identificación de variables que afectan el esfuerzo para la creación de soluciones, caracterización de tipos de proyectos de acuerdo a las variables asociadas

y finamente construcción de una herramienta de simulación y pruebas.

A continuación se presenta el diagrama de flujo que describe de forma global modelo para calcular el esfuerzo en etapas tempranas:

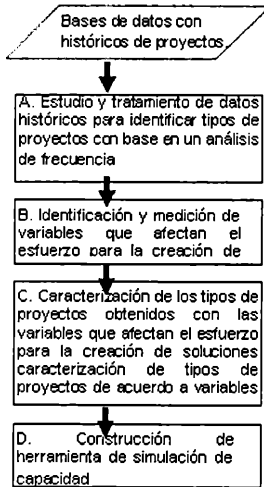


Figura 1: Diagrama de flujo con el procedimiento para la creación del modelo dinámico de estimación temprana

A Estudio y Tratamiento de Datos Históricos

En esta actividad se hizo un estudio estadístico para determinar de qué tipo de distribución provenían los datos relacionados con el esfuerzo (horas-hombre) requerido en el desarrollo de proyectos, para luego dar una posible clasificación de proyectos.

El estudio se orientó a buscar la normalidad de los datos, o la forma de obtenerla, teniendo en cuenta que existe más información acerca del tratamiento de datos que siguen una distribución normal.

En la Gerencia Informática, los datos del esfuerzo se mueven en un rango bastante amplio de 0,25 a 15620 horas-hombre. Por medio de un histograma de frecuencias (Figura 2) y algunas pruebas de normalidad (Prueba de Shapiro-Wilk y prueba de Anderson-Darling), se pudo determinar que los datos no seguían una distribución normal.

Para ajustar los datos a una distribución normal, se aplicó logaritmo natural y se eliminaron los proyectos que tienen un esfuerzo menor de 5 horas-hombre y mayor de 994.858 horas-hombre. Posteriormente, usando pruebas de bondad de ajuste, reincorporando los datos que habían sido excluidos, y ha-

ciendo otros procedimientos estadísticamente válidos, se obtuvieron 13 intervalos tal como se muestra en la Tabla 1.

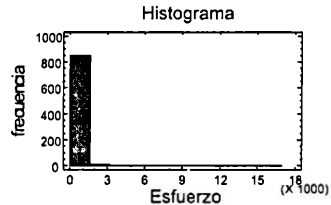


Figura 2: Histograma de frecuencia del esfuerzo (horas-hombre) requerido para el desarrollo de proyectos software

Tabla 1: Clasificación de proyectos de acuerdo a su esfuerzo asociado

Intervalo	Límite Inferior	Límite Superior
T0	0,25	4,9998
T1	4,9998	7,2087
T2	7,2087	13,0771
T3	13,0771	19,9846
T4	19,9846	28,5604
T5	28,5604	39,6226
T6	39,6226	54,4967
T7	54,4967	75,6018
T8	75,6018	108,0485
T9	108,0485	165,1229
T10	165,1229	299,5406
T11	299,5406	994,8580
Tx	994,858	15620

La Tabla 1 muestra la clasificación de proyectos asociados a la creación de soluciones en la Gerencia Informática de Orbital de acuerdo al esfuerzo requerido para su desarrollo.

El intervalo T0 corresponde a los proyectos que están entre 0,25 y 5 horas-esfuerzo. Los intervalos T1 a T11 fueron obtenidos con el test de bondad de ajuste, y, el intervalo Tx corresponde a los proyectos que superan las 994,858 horas-hombre. Si bien los extremos de los intervalos obtenidos con los históricos disponibles son bastante distantes, el modelo creado y la herramienta de simulación, permiten que los intervalos se ajusten de manera dinámica a medida que se ingresan nuevos proyectos.

En apartados posteriores, los tipos de proyectos obtenidos serán caracterizados con variables obtenidas para tal fin.

B Identificación y medición de variables que afectan al esfuerzo En esta actividad se estableció un método para descubrir las variables, eventos y retardos que aumentan el esfuerzo en el proceso de creación de soluciones, incluyendo para cada una de ellas: *nombre, descripción, unidades de medida, clasificación de valores y comentarios*. Dichas variables se clasificaron en "Variables generales" y "Variables por equipo". Las variables generales son aquellas que después del análisis realizado, aplican para todos los equipos de la Gerencia y las variables por equipo son variables que aplican a equipos particulares. A continuación se presentan algunas variables y sus atributos.

Variables generales

Para la evaluación y medición de variables como "conocimiento del negocio", "conocimiento de nuevas tecnologías" y "conocimiento técnico", las cuales hacen parte del conjunto de variables generales, se construyó un método, el cual no es ampliado en este trabajo.

Variables por equipo

El resultado de esta actividad fue utilizado para caracterizar los tipos de proyectos. La combinación de las variables asociadas a una nueva solución, permite determinar la complejidad de una solución en términos del esfuerzo requerido.

C Caracterización de tipos de proyectos Luego de tener los resultados de las actividades de identificación de tipos e identificación y medición de variables, se realizó una caracterización de los tipos de proyectos con la intervención de analistas y directores de la Gerencia de Informática de Orbitel. Para el estudio se tomó una muestra aleatoria de proyectos incluidos en cada uno de los intervalos de esfuerzo (tipos de proyectos) obtenidos con el análisis estadístico.

En plantilla mostrada en la Figura 3, los proyectos representan la muestra de proyectos seleccionada tomando como base el total de proyectos que cada equipo realizó desde enero del año 2003 hasta marzo del año 2005. Las variables obtenidas en la segunda actividad, las cuales fueron asociadas a cada uno de los tipos de proyectos con base en la experiencia de los analistas de la Gerencia Informática y en la información recopilada de cada proyecto.

Los resultados obtenidos en esta actividad, permiten que los analistas sólo necesiten indicar las variables que intervendrán en la nueva solución y el modelo las combinará hasta conseguir el tipo que mas se ajusta al nuevo proyecto. Es importante anotar que el modelo permite la inclusión de nuevas

variables y tiene la capacidad de refinarse dinámicamente con cada nuevo proyecto incluido en la base de datos.

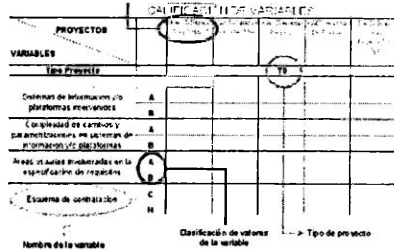


Figura 3: Plantilla de asociación de variables a tipos de proyectos

D Implementación de la aplicación Luego de hacer el tratamiento estadístico de los datos y tener una clasificación de proyectos de acuerdo a las variables asociadas a cada tipo, fue necesario diseñar e implementar un algoritmo que le permitiera al analista conocer el tipo de proyecto a gestionar. Luego de seleccionar las variables involucradas.

También fue necesario construir una base de datos para almacenar la clasificación de proyectos, las variables relacionadas, y otra información relevante. Todo lo anterior se integró para obtener una aplicación orientada a la web que permite a los analistas monitorear su capacidad, con sólo ingresar los proyectos que desean gestionar.

A continuación se muestran algunas pantallas generadas por la aplicación y sus respectivos comentarios:

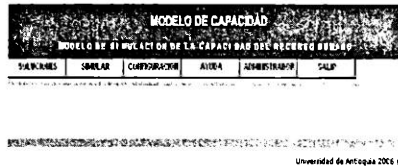


Figura 4: Barra de menús de la aplicación

Los dos menús más relevantes de la aplicación son el menú "soluciones" y el menú "Simular". El menú *SOLUCIONES* permite ingresar una nueva solución, asignar participación de un analista y modificar una solución existente.

En el menú *SIMULAR* se pueden obtener series de tiempo sobre la capacidad empleada de un analista, un equipo, una dirección, y la Gerencia Informática. Los

resultados obtenidos se mostrarán en la sección de resultados.

Para ingresar una nueva solución o proyecto en el cual intervendrá un analista se utiliza la interfaz mostrada en la Figura 5.

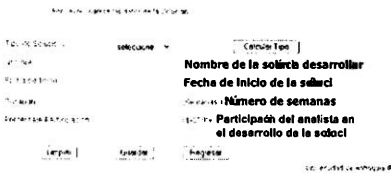


Figura 5: Interfaz para ingreso de nuevas soluciones

En este formulario, el analista debe ingresar datos como: nombre del proyecto o solución, fecha de inicio, duración, porcentaje de participación y tipo de solución (intervalo de esfuerzo) en caso de conocerlo. Si se trata de una estimación temprana, la herramienta de simulación permite calcular el tipo de proyecto de acuerdo a las variables involucradas (Figura 6)



Figura 6: Herramienta para calcular el tipo de solución

El formulario mostrado en la Figura 6 permite señalar cuales variables (con sus respectivos niveles) están presentes en la solución que gestionará el analista, luego, la herramienta simula se encarga de indicar en qué nivel de la clasificación se encuentra una solución en la que intervienen las variables ya señaladas por el analista

4 RESULTADOS DEL "MODELO DE CAPACIDAD"

En la actualidad, el resultado de este proyecto de investigación aplicada es un modelo que le permite a los analistas de la Gerencia Informática no solamente monitorear su capacidad y determinar cuando están en situación de sobreesfuerzo, sino que también permite hacer estimaciones tempranas con base en históricos y experiencias

de proyectos similares.

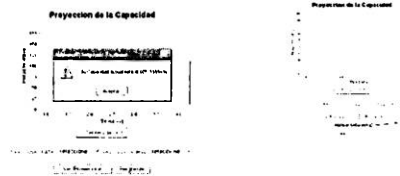


Figura 7: Interfaz gráfica del modelo de capacidad

La interfaz del modelo esta diseñada de manera que si la capacidad sobrepasa el 100%, al simular algún mes, se presentará una señal de alerta, buscando que el analista ajuste su capacidad y no entre en condiciones de sobre-esfuerzo (Figura 7). La alerta seguirá mostrándose hasta que el analista haga los ajustes necesarios.

Además, en la simulación es posible observar el comportamiento de la capacidad al agregar una solución de un tipo determinado, sin necesidad de almacenar esta solución en la base de datos. Como lo muestra la Figura 7, en la parte inferior aparecerá un listado de las soluciones asociadas al analista.

El software permite también mostrar gráficos del comportamiento de la capacidad por equipo, por Dirección, y de toda la Gerencia Informática, informando siempre si se ha entrado en condición de sobre-esfuerzo.

5 CONCLUSIONES

La simulación de procesos de gestión y desarrollo software relacionados con tareas de estimación temprana, representan un importante avance para entender, entre otros, el impacto del esfuerzo asociado a la creación de soluciones sobre las tareas de operación y de atención de fallas.

La creación de bases de datos con acumulación dinámica de históricos de proyectos permitió establecer de manera muy aproximada el conjunto de variables y funciones caracterizan los procesos de creación de soluciones de software.

El modelo dinámico usa un número de parámetros, variables y funciones que caracterizan los proyectos y el entorno de una Gerencia Informática. Para realizar una estimación temprana de un nuevo proyecto, es necesario seleccionar las variables que caracterizan la nueva solución, una vez se cuenta con el grupo de variables que caracterizan la nueva solución, el modelo puede además de realizar la estimación temprana de esfuerzo, planear y simular la capacidad del recurso humano.

Considerando que las variables identificadas en esta investigación son frecuentes para cualquier gerencia informática, debido a que se establecieron para los proce-

sos sugeridos por la metodología ITIL y que la base de datos con históricos de proyectos puede crecer de manera dinámica, el modelo creado es una herramienta que puede ser usada para hacer estimaciones y experimentaciones cuando se requiere hacer una estimación con una especificación de requisitos insuficiente y cuando la información disponible en bases de datos con históricos de proyectos sea incompleta.

REFERENCIAS

- Albrecht, A. (1979), Measuring application development productivity, in 'Proceedings SHARE/GUIDE IBM Applications Development Symposium, Monterrey, CA'.
- Albrecht, A. y Gaffney, J. (1983), 'Software function, source lines of code, and development effort prediction: A software science validation', *IEEE Transactions on Software Engineering* SE-9(6), 639-648.
- Barbacci, J. (1985), 'The software engineering institute: bridging practice and potential', *IEEE software* pp. 4-21.
- Jones, C. (1999), 'Activity-based software costing', *Computer* pp. 103-101.
- Jorgensen, M. y Sjøberg, D. (2001), 'Impact of effort estimates on software project work', *Information and software technology* 43, 939-948.
- Peralta, M. (2004), 'Estimación del esfuerzo basada en casos de uso', *Reportes Técnicos en Ingeniería de Software. Buenos Aires - Argentina* 6(1), 1-16.
- Ribu, K. (2005), Estimating object-oriented software projects with use cases, En línea: <<http://heim.ifi.uio.no/kribu/oppgave.pdf>> c10'05, Universidad de Oslo.
- Sánchez, F. (1999), Medida del tamaño funcional de aplicaciones software, Technical report, Universidad de Castilla - La Mancha.
- Toro, M. (2002), 'Marco dinámico integrado para la mejora de los procesos software'.
- Varas, M. (2002), 'Una experiencia con la estimación del tamaño del software', En línea: <<http://www.inf.udec.cl/revista/edicion1/nvaras.htm>>.
- Zhenyou, J. (2004), Reduced models of software development effort estimation, Technical report, Department of electrical and computer engineering. University of Alberta. Edmonton, Alberta.