

PROPUESTA METODOLOGICA PARA LA EDUCIÓN DE REQUISITOS EN PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Paola Britos^{1,2}, Enrique Fernandez^{1,2}, Ramón García-Martínez^{1,2}
Centro de Ingeniería del Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA.
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA
{pbritos, rgm}@itba.edu.ar

RESUMEN

Existen diversas metodologías de explotación de información que destacan la importancia de llevar adelante una educación de requisitos ordenada, documentada, consistente y trazable a lo largo de todo el proyecto. Como antecedente a este tipo de documentación podemos citar el Standard 830 de IEEE en el que se propone un esquema de documentación de requisitos para proyectos de Ingeniería del Software. Dicho esquema no se adapta en su totalidad a proyecto de explotación de información, es por eso que es objetivo de este artículo es plantear un modelo de documento de requisitos para proyectos de explotación de información que sea consistente, no ambiguo y trazable a lo largo de todo un proyecto de explotación de datos.

1. Introducción

Varias metodologías de explotación de información [Pyle, 2003; Chapman *et al.*, 2000] coinciden en señalar como una de las actividades tempranas del proyecto la especificación de requerimientos. Sin embargo ninguna de ellas identifica técnicas para educir el conocimiento necesario ni propone plantillas para la documentación sistemática del relevamiento en cuestión.

El Standard 830 del [IEEE, 1993], y en SWEBOK Guide [IEEE Computer Society, 2004] formulan recomendaciones sobre prácticas a seguir para la especificación de requerimiento de software. El Standard 830 del IEEE [3] proponen plantillas para documentar la especificación de los requerimientos relevados así como criterios para asegurar la correcta redacción de los mismos.

2. Descripción del problema

El Standard IEEE 830 estructura el informe de requerimientos en tres siguientes componentes: introducción, descripciones generales, especificación de requisitos. La introducción comprende la educación del objetivo del proyecto software, definiciones, acrónimos, abreviaturas y referencias. En el contexto de proyectos de explotación de información surge la necesidad de educir criterios de éxito del proyecto. En las descripciones generales se presentan los resultados de educir: la perspectiva del objeto tecnológico a producir, las características de los usuarios, las restricciones, suposiciones y dependencias. En el contexto de proyectos de explotación de información surge la necesidad de educir los riesgos y contingencias, los costos y beneficios y la evaluación de herramientas y técnicas. En la especificación de requisitos, por cada requisito se debe educir: su objetivo, el origen de la información, los supuestos, las restricciones, los atributos involucrados y cualquier información de soporte.

Este Standard no satisface completamente las características de proyectos de exploración de información. A continuación se presenta un análisis de las limitaciones de dicho Standard:

- Apto para proyectos de exploración de información: los aspectos de definición del propósito general de la organización, definición del objetivo que persigue la organización, definición de terminología utilizada, especificación de las restricciones, suposiciones y dependencias.

- Necesitan modificación para proyectos de exploración de información Deben ser adaptados: la definición de la perspectiva de producto, la definición de las características de los usuarios, la especificación de los requerimientos (interfases, funcionalidades, restricciones de equipamiento, atributos).
- No contemplado para proyectos de exploración de información: definición de potenciales riesgos y contingencias, definición de costos y beneficios, definición de criterios de éxito, siendo que todos estos aspectos influyen en la especificación de los requerimientos.

En este contexto es un problema a resolver la creación de un Standard para proyectos de exploración de información.

3 Solución propuesta

A continuación se presenta una propuesta de Standard para la definición de requisitos para proyectos de exploración de información.

3.1. Características de una buena especificación de requisitos para proyectos de explotación de información

La definición de requisitos de un proyecto debe:

- Ser correcta: No se debe especificar requisitos erróneos.
- No ambiguo: No se deben especificar requisitos que se presten a varias interpretaciones.
- Completo: Cada requisito debe ser citado de forma completa. Esto implica que todos los requisitos deben ser significativos a los usuarios del proyecto.
- Consistente: Cada requisito no debe contradecirse con los definidos anterior o posteriormente.
- Verificable: Debe poder corroborarse cada requisito con algún representante de la organización.
- Trazable: Debe quedar documentado y con guías para conocer el estado del requisito en los diversos avances del proyecto.

3.2. Componentes del documento de educación de requisitos

En este ítem se indica las partes que debe contener del documento de educación de requisitos para proyectos de explotación de información.

1. Introducción
 - 1.1. Objetivos del proyecto
 - 1.2. Definiciones, acrónimos y abreviaturas
 - 1.3. Criterios de éxito del proyecto
2. Descripciones generales
 - 2.1. Expectativa del proyecto
 - 2.2. Recursos humanos involucrados
 - 2.3 Restricciones
 - 2.4. Suposiciones y dependencias
 - 2.5. Riesgos y contingencias
 - 2.6. Costos y beneficios
 - 2.7. Evaluación de herramientas

3. Requisitos

- 3.1. Objetivo del requisito
- 3.2. Origen de la información del requisito
- 3.3. Supuestos del requisito
- 3.4. Restricciones del requisito
- 3.5. Atributos involucrados en el requisito
- 3.6. Información de soporte sobre el requisito

Las etapas 3.1 a 3.6 se repetirán tantas veces como requisitos identificados.

3.2.1. Introducción (sección 1)

El propósito de este ítem es proveer una visión general del proyecto, la misma contiene:

Objetivo del proyecto: En esta sección se debe describir cual es el objetivo del proyecto de explotación de información, describir la motivación que lleva a la realización del mismo.

Definiciones, acrónimos y abreviaturas: Esta sección describe los términos, acrónimos y abreviaturas utilizadas en el proyecto; obteniéndose así un vocabulario común entre los recursos humanos involucrados en el proyecto.

Criterios de éxito del proyecto: En esta sección se debe describir cuales con los criterios por los cuales el proyecto se considera que es exitoso, debe ir en relación con los objetivos planteados en el primer ítem de esta sección (objetivo del proyecto).

3.2.2. Descripciones generales (sección 2)

El propósito de este ítem es proveer una explicación general de proyecto, en la misma no se especifican los requisitos sino que provee el entorno en el que los mismos se encuentran involucrados. Esta sección contiene:

Expectativas del proyecto: En la cual se describe que se espera del proyecto de explotación de información. Las expectativas descritas deben estar en coincidencia con el objetivo del proyecto (ítem 1) y los criterios de éxitos descriptos (ítem 3) en la sección 1.

Recursos humanos involucrados: Se menciona los recursos involucrados en el proyecto y su rol en el proyecto. Debe incluir los recursos involucrados como expertos en exploración de proyectos como también los expertos en el dominio de negocio sobre el cual están involucrados los datos.

Restricciones: Las restricción indicaran cuales son los limites a priori que se conocen sobre el proyecto, pueden ser:

- [a] de organización: política (por ejemplo: oportunidad de mercado para lanzamiento de un producto) y legales (por ejemplo: normativa de habeas data en cuanto a los datos personales de los clientes);
- [b] de datos: acceso a las fuentes de información (por ejemplo: los datos de los movimientos de tarjetas de créditos son privados), calidad de datos (por ejemplo: los campos que contienen localidad se encuentran en un alto porcentaje en blanco) y cantidad de datos (por ejemplo: solo se tiene datos 2 meses con poca cantidad de información);
- [c] de recursos humanos y técnicos: tamaño de las fuente de datos en relación al hardware y software (por el tamaño de la base de datos y recursos de software y hardware con que se cuenta el procesamiento y unificación de los mismos se hace de forma distribuida), limitaciones de hardware (por ejemplo: No se cuentan los datos en un servidor propio, sino que es compartido por varias aplicaciones), limitaciones de software (por ejemplo: El software que contiene las fuentes de información no es compatible entre ellos) y recursos humanos (por ejemplo: Los

recursos tienen mucha antigüedad por lo que le corresponden muchos días de vacaciones y el proyecto se encuentra en ejecución en momento de toma de vacaciones); y

[d] de proyecto: actividades que afectan al proyecto (por ejemplo actualización permanente de los datos) y seguridad (por ejemplo el acceso determinada documentación que es parte del proyecto solo se puede hacer en un lugar determinado de la organización sin posibilidad de tener copia de la misma).

Suposiciones y dependencias: En este ítem se describirán los supuestos que se tiene del proyecto, así como las dependencias que tenga el proyecto con otros proyectos de la organización.

Riesgos y contingencias: En este ítem se debe incluir la gestión de riesgo de un proyecto de explotación de información. La gestión de riesgo en este tipo de proyectos consiste en el análisis de la organización a través de métodos, y herramientas para la gestión. La misma mantiene un ambiente disciplinado proactivo que evalúa lo que puede salir mal continuamente; determina qué riesgos son importantes; y las acciones para tratar con esos riesgos.

Costos y beneficios: En este ítem se indicaran los beneficios de la realización del proyecto así como los costos involucrados, el mismo deberá ser consistente con los criterios de éxitos definidos en la sección 1 - ítem 3, así como también con las expectativas citadas en el ítem 1 de esta sección.

Evaluación de herramientas: En este ítem se debe indicar la importancia del impacto de una herramienta de explotación de información en la organización y la inversión que la misma debe hacer en términos económicos; también hacen que el proceso de selección de las mismas para lo cual es conveniente realizar una análisis cualitativo y cuantitativo de cada una de ellas, para que la organización pueda escoger la que mejor se adapte a sus necesidades.

3.2.1. Requisitos (sección 3)

El propósito de esta sección es proveer los requisitos necesarios para poder cumplir con los objetivos previstos en el ítem 1 de la sección 1 del documento de especificación de requisitos para proyectos de explotación de información. Se debe destacar que por cada requisito se deben identificar los ítems 1 a 6 que se describen a continuación:

Objetivo del requisito: En este ítem se debe incluir el objetivo del requisito en función lo especificado en el ítem 1 - sección 1 (objetivos del proyecto) y justificar porque se considera que este requisito ayuda a cumplir con el objetivo citado.

Origen de la información del requisito: Es conveniente que en este punto se indique toda la información que indique de donde provienen los datos para poder cumplir con el requisito, así como los responsables de dichos datos.

Supuestos del requisito: En este ítem se debe indicar las suposiciones que se tiene sobre los resultados de este requisito, dichas suposiciones deben ser consistentes con el objetivo del proyecto (ítem 1, sección 1), las expectativas del mismo (ítem 1, sección 2) y las suposiciones y dependencias (ítem 2, sección 4).

Restricciones del requisito: En este ítem se debe indicar las restricciones que se tiene sobre este requisito, dichas restricciones deben ser consistentes las restricciones planteadas en la sección anterior (ítem 3, sección 2).

Atributos involucrados en el requisito: En este punto se debe indicar los atributos que se presuponen útiles para cumplir con el objetivo del requisito, como así también su origen.

Información de soporte para el requisito: Este ítem tiene por objetivo identificar todo tipo de información que sirva de referencia y apoyo para cumplir con el objetivo del requisito en cuestión.

4 Conclusiones

Se puede observar que a través del Standard propuesto para proyectos de explotación de información se lleve adelante una especificación de requisitos adecuada, consistente y trazable a lo largo de todo el proyecto. Esto permite una gestión ordenada del proyecto lo que garantiza un entendimiento adecuado de los requisitos solicitados en este tipo de proyecto. Como futuras líneas de investigación se propone:

- [a] Formulación de un ciclo de vida para la educación de requisitos de un proyecto de explotación de información.
- [b] La generación de plantillas modelos para la educación de requisitos de acuerdo al Standard planteado para las distintas fases de un proyecto de explotación de información.
- [b] Formular técnicas de educación propias de la explotación de información.

5. Formación de Recursos Humanos.

A la fecha de esta comunicación se han radicado en el Proyecto: una tesis doctoral, dos tesis de magister, cuatro tesis de grado en ingeniería.

6. Referencias

- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. *CRISP-DM 1.0 Step by step data mining guide*. Editado por SPSS. <http://www.crisp-dm.org/CRISPWP-0800.pdf>. Pagina vigente al 14/09/06.
- IEEE 1993. *Standard IEEE 830-1993: Recommended Practice for Software Requirements Specifications*. Institute of Electronic and Electrical Engineers Press.
- IEEE Computer Society. 2004. *Guide to the Software Engineering Body of Knowledge*. ISBN 0-7695-2330-7. <http://www.swebok.org> (vigente a diciembre 2006)
- Pyle, D. 2003. *Business Modeling and Data Mining*. Morgan Kaufmann. ISBN 1-55860-653-X.