

Calidad de Datos como Valor Estratégico de la Información en E-Business

Marta E. Cabrera Villafañe, Germán Antonio Montejano, Mario Berón
Departamento de Informática
Universidad Nacional de San Luis
San Luis - Argentina
email: cabreramarta@arnet.com.ar
email: {gmonte,mberong@unsl.edu.ar }

Resumen

Uno de los factores esenciales, para que una empresa se posicione en altos niveles de competitividad en el mercado, es el acceso rápido, eficiente y fácil a información valiosa y útil.

Para alcanzar este objetivo, es necesario que los datos que se obtienen a partir del funcionamiento de la empresa sean de Calidad. Calidad es una característica fundamental que debe poseer los datos que son utilizados como materia prima por los procesos de producción de información. La calidad de la información es una particularidad fundamental que debe tener un producto de información para que su utilización cumpla con los requisitos de los usuarios.

En este artículo se describe una línea de investigación centrada en la definición de las características de calidad que deben poseer los datos de una empresa para que la misma se posicione en los primeros niveles de competitividad.

1. Contexto

Esta línea de investigación se desarrolla en el contexto del proyecto de investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de Ingeniería de Software en Evolución. Dicho proyecto se desarrolla en la Universidad Nacional de San Luis y es la continuación de proyectos de investigación anteriores que, al igual que el presente proyecto, han sido aprobados por el programa de incentivos.

El proyecto mencionado en el párrafo precedente posee una amplia trayectoria en el desarrollo de trabajos de investigación a nivel nacional e internacional. Dicha trayectoria está avalada por la realización de varias tesis de doctorado y maestrías como así también en la realización de publicaciones en conferencias de reconocida importancia.

2. Introducción

En la actualidad es posible encontrar gran variedad de software de Business Intelligence con aplicaciones similares que pueden ser utilizadas en las diferentes áreas de la empresa, como por ejemplo: ventas, marketing, finanzas, etc. Son numerosas las empresas que se han beneficiado por la implementación de un sistema de Business Intelligence, pronosticándose que con el tiempo se convertirían en una necesidad para toda empresa. La Internet y todos los servicios que ofrece proporcionan un sin número de datos que si son bien utilizados forman un patrimonio importante para el progreso económico de la empresa. Según los gerentes y personal encargado de la toma de decisiones, una empresa mantendrá su nivel de competitividad si los clientes tienen un acceso rápido y fácil a información útil y valiosa brindada por la empresa. Una forma de alcanzar este objetivo consiste en proporcionar datos, obtenidos con el uso de Business Intelligence, con alta calidad. Según Tomas Siebel, el e-business, no se centra en la Web sino en el cliente. Por esta razón, las empresas implementan técnicas de E-Business que

permiten a la organización: Registrar, Medir, Actualizar y Analizar, gran cantidad de información relacionada con el cliente [ISO07]. Además las empresas pueden añadir a sus bases de datos de cliente datos externos, recuperados por los servicios de información, que permitan comprender mejor las necesidades de los mismos. La línea de investigación presentada en este artículo tiene como objetivo principal brindar una solución al siguiente problema: "Calidad de Datos como valor estratégico de la Información en E-Business". Para alcanzar este objetivo se piensa que la definición de una Ontología de Medición de la Calidad de Datos es una tarea fundamental [GGOCM03, Mc01, AdM09]. Se puede hacer esta afirmación porque una ontología permite: i) Describir sin ambigüedades un dominio de discurso y ii) Establecer un vocabulario, compartido y consensuado. Una vez realizada esta tarea se procederá a la definición de métricas de calidad [Zu98, KBHSL01, OI00].

Calidad es una característica fundamental que debe poseer los datos que son utilizados como materia prima por los procesos de producción de información. La calidad de la información es una particularidad fundamental que debe tener un producto de Información para que su utilización cumpla con los requisitos de los Usuarios. El artículo está organizado como sigue. La sección 3 describe la importancia de los datos de calidad para e-business. La sección 4 presenta resultados obtenidos referentes a la definición de una ontología de calidad de datos en la Web semántica. Finalmente la sección 5 describe las tareas realizadas en el contexto de la formación de recursos humanos.

3. Calidad de Datos como Valor Estratégico en E-Business

Uno de los principales propósitos de esta línea de investigación es construir una ontología que permita medir la calidad de los datos.

Esta tarea tiene como finalidad proveer las bases conceptuales para desarrollar métricas válidas para esta dimensión.

Crear una Ontología no es una tarea sencilla porque implica tener en cuenta los aspectos que se detallan a continuación:

- Identificar el propósito de la ontología: El objetivo de esta etapa es dejar claro por qué se construye la ontología y qué uso se le pretende dar. También es útil en este momento identificar y caracterizar el rango de posibles usuarios y aplicaciones de la ontología.
- Construir la Ontología: En esta etapa se distinguen tres sub-tareas, ellas son: Capturar, Evaluar y Documentar la Ontología. La primera está relacionada con: i) Identificar los conceptos claves y relaciones en el dominio de interés, ii) La producción de definiciones textuales precisas y carentes de ambigüedad que describan tales conceptos y relaciones y iii) Identificar los términos para referirse a tales conceptos y relaciones. La segunda se centra en evaluar los ambientes de software asociados y la documentación con respecto al marco de referencia relacionado con la ontología [ISO08, ISO99]. Finalmente, la tercera estudia el establecimiento de pautas para documentar la ontología construida teniendo presente el propósito de la ontología).

Para llevar a cabo las tareas mencionadas previamente se realizaron dos actividades muy importantes. La primera de ellas fue la selección de la metodología para la construcción de la ontología, dicha tarea se explica en la subsección siguiente. La segunda, descrita en la sección 4, consistió en la aplicación de dicha metodología para desarrollar la ontología en cuestión.

3.1. Methontology: Una Metodología para el Desarrollo de Ontologías

Para la construcción de la ontología se seleccionó la metodología METHONTOLOGY, por ser una de las metodologías más difundidas, maduras que cuenta con actividades y técnicas para el proceso de desarrollo de ontologías.

La ontología resultante del proceso de definición será aplicada al diseño y construcción de una ontología de medición de la calidad de los datos basado en la Web Semántica [MO01, OI00, APJ08], que permitirá explorar, reusar y compartir información relacionada a métricas. Los principales pasos propuestos por esta metodología son:

- Paso 1: Establecimiento del Proceso de Desarrollo de la Ontología: En esta etapa se identifican que actividades se realizan al construir la ontología. Se deben Definir las Actividades del Proyecto (incluyen planificación, control y aseguramiento de la calidad), las Actividades Orientadas al Desarrollo del Proyecto (incluyen especificación, conceptualización, formalización, implementación y mantenimiento) y las Actividades de Apoyo que incluyen una serie de actividades ejecutadas al mismo tiempo que las actividades orientadas al desarrollo, sin las cuales la ontología podría no llegar a ser construida, (incluyen adquisición de conocimiento, evaluación, integración, documentación y administración de versiones).

- Paso 2: Desarrollo de la Ontología: El ciclo de vida para el desarrollo de una ontología propuesto por Methontology incluyen las siguientes actividades orientadas al desarrollo:

1. Especificación: El propósito de la especificación es definir en un documento cuál es el principal objetivo de la ontología, con que propósito se desarrolla (posibles aplicaciones), cuál es su nivel de generalidad y granularidad, y cuál es su alcance (establecimiento de los límites a cubrir).

2. Conceptualización: Estructura el conocimiento del dominio utilizando modelos conceptuales en un nivel de conocimiento. Una vez que se ha adquirido el conocimiento del dominio en cuestión, los desarrolladores tienen un conjunto de conocimiento no estructurado que debe ser organizado. La actividad de conceptualización tiene como objetivo organizar y estructurar el conocimiento adquirido, usando un conjunto de Representaciones Intermedias (IR), que son

modelos para representar conceptos independientes de los lenguajes y ambiente de desarrollo que se usaran en la implementación de la ontología. Gómez Pérez y otros, en su trabajo "Towards a Method to Conceptualize Domain Ontologies", propone un conjunto de presentaciones intermedias como (diccionario de datos, árboles de clasificación, tablas de atributos entre otros), para la conceptualización de ontologías de dominio.

3. Formalización: Transforma el modelo conceptual en un modelo formal o semi-formal computable.

4. Implementación: Se construyen modelos computables en un lenguaje computacional como podría ser Ontolingua, RDF/S (Resource Description Framework/Schema) u OWL (Ontology Web Language).

5. Mantenimiento: Actualiza y corrige la ontología.

4. Resultados y Objetivos

Los resultados obtenidos en la construcción de la ontología son los siguientes:

- Se definió un glosario de conceptos, en este contexto se entiende por concepto a las ideas básicas que se intentan formalizar. Los conceptos pueden ser: clase de objeto, métodos, planes, estrategias, proceso de razonamientos, etc. Como resultado de la realización de este paso se pudieron identificar los conceptos relevantes para el tema de estudio. Algunos de ellos son: Atributo, Concepto Calculable, Criterios de Decisión, Escala Categórica, etc.

- Se precisaron las interrelaciones semánticas: Esto se logró de la siguiente manera: 1) Definiendo los conceptos y las relaciones existentes entre ellos y representándolos con una herramienta o lenguaje estándar. Por ejemplo, se podría utilizar un diagrama de clases del lenguaje UML. 2) Elaborando una tabla de clases y relaciones con las diferentes clases que se van identificando.

- Se establecieron las siguientes relaciones: i) Asociación con: Uno o más atributos medibles se asocian con una o más entidades; ii) Automatizados por: Uno o más métodos pueden ser automatizados por varias o ninguna herramienta de software. iii) Asociaciones: Un concepto calculable asocia (combina) uno o más atributos medibles, iv) Contenido: Una métrica o un indicador contienen una escala específica, v) Describe: Una o más conceptos calculables se definen para satisfacer una necesidad de información correcta. De esta manera, un concepto calculable describe una necesidad de información concreta, vi) Cuantifica: Una o más métricas cuantifican un atributo; vii) Origina: Una actividad de medición (o cálculo) produce un valor de medida (o indicador) específico.
- Se identifican los atributos: Para cada concepto se identifican sus atributos. Dichos atributos son especificados en una tabla de atributos e incluidos en la especificación de la ontología. Los atributos identificados son: i) Concepto o Atributo: Nombre, Definición, Objetivo; ii) Concepto Calculable: Nombre, Descripción; iii) Cálculo: Intervalo, iv) Escala Categórica: etc.
- Se construyó una tabla de conceptos: En esta tabla se especificaron los atributos de cada concepto y se los incluye en la especificación de la ontología (diagrama de clases de UML).
- Se definió un conjunto de Axiomas, con la finalidad de proveer conocimiento adicional. Estos axiomas expresan otras relaciones entre los conceptos y restringe su interpretación, agregando restricciones estructurales al diagrama conceptual. Los axiomas son expresados en lenguaje natural y en lógica de primer orden, donde los términos de la ontología se expresaron a través de predicados unarios o binarios, según sean conceptos o relaciones respectivamente.
- Se implementó la ontología en lenguaje RDFS, por ser uno de los lenguajes estándar recomendado por el W3c, con la aplicación del editor de ontología PROTÉGÉ.

El principal resultado de esta investigación consistió en la definición de una ontología para el ámbito de la calidad del Dato en la Web Semántica ya que provee una base conceptual para compartir, reusar y organizar toda la información relacionada a la calidad del dato, en el área de medición y evaluación de software, de una forma extensible, interoperable y automáticamente procesable.

Las posibles extensiones, que establecen nuevas líneas abiertas que pueden ser objeto de desarrollo en un futuro próximo son:

- Construir una ontología que permita la elaboración de representación intermedia a través de gráficos y tablas. El propósito de este trabajo es facilitar la comprensión de expertos en el dominio con escaso marco conceptual lo que a ingeniería ontológica se refiere. Esta aproximación posibilitará un enriquecimiento terminológico en cada una de las interacciones.
- Desarrollar una notación gráfica basada en UML que permita representar modelos de medición y posteriormente, la implementación de una herramienta de definición de modelos de medición que dé soporte a dicha notación.
- Utilizar/elaborar una metodología para la clasificación automática de documento XML según su calidad de datos asociada usando principios de soft computing (término utilizado que engloba diversas técnicas empleadas para solucionar problemas que manipulen información incompleta, con incertidumbre e inexacta.

5. Formación de Recursos Humanos

Los trabajos realizados en esta línea de investigación están siendo desarrollados como parte de una tesis de maestría en Ingeniería de Software en la Universidad Nacional de San Luis. Dicho trabajo está en la etapa de finalización y dentro de poco tiempo será sometido a evaluación por un jurado pertinente. Se espera en futuro continuar con la investigación en este campo con el propósito de generar

otras tesis de maestría como así también de doctorado.

Referencias

[MO01]

Martín , María de los A; Olsina ,Luis. Fortalezas y Limitaciones de las Tecnologías Actuales para el Procesamiento Semántico en la Web. GIDIS, Departamento de Informática, UNL Para Semántica Web. <http://www.w3.org/2001/sw/>. 2001.

[Zu98] Zuse , Horst Walter. A Framework of Software Measurement. Gruyter - ISBN 3-11-015587-

7. Berlin, New York . 1998.

[KBHSL01] Kitchenhan, Barbara A.; Hughes, Robert; Stephen, T; Linkman, G. Modeling software Measurement Data. ACM Digital Library. Vol: 27 Issue: 9. New York , EEUU. Septiembre 2001.

[GGOCM03] Genero , M; Garcia , F; Olsina ,L.; Calero, C; Martín , M. An Ontology for Software Measurement. 15 th International Conference on Software Engineering and Knowledge Engineering. ISBN: 1-891706-12-8, pp.307-317. 2003.

[OI00] Olsina, Luis. Quantitative Methodology for Evaluation and comparison of Web Site Quality. Doctoral Thesis. Facultad de Ciencias Exactas, Universidad Nacional de La Plata. Argentina. 2000.

[ISO07] El estándar ISO/IEC 15939

[ISO15939].ISO(International Organization for Standardization)/ IEC Comisión Electrotécnica Internacional Comisión). Suiza. 2007.

[ISO08] La Calidad del Producto { Parte 1 Modelo de Calidad. ISO/IEC 9126-1. Nueva Zelanda. Australia. 2008.

[ISO99] El estándar ISO/IEC 14598-1

[ISO14598]. ISO (International Organization for Standardization)/ IEC Comisión Electrotécnica Internacional Comisión. Suiza. 1999.

[ISO08] La Calidad del Producto { Parte 1 Modelo de Calidad. ISO/IEC 9126-1. Nueva Zelanda. Australia. 2008.

[ISO99] El estándar ISO/IEC 14598-

1[ISO14598]. ISO (International Organization for Standardization)/ IEC. Comisión Electrotécnica Internacional Comisión. Suiza. 1999.

[Mc01] McGuinness, Deborah L. \Ontology development 101", Una guía para crear

su primer Ontología. <http://www.ksl.stanford.edu/software/ontolingua/>. Páginas: 1-23. 2001.

[APJ08] Avancha, Sasikanth; Patan, Chintan; Joshi , Anupam. Ubiquitous and Pervasive Applications. Primera Conferencia Anual Internacional sobre sistemas Móviles y computación Ubicua. ISBN 0-7695-2208-4. 2008.

[AdM09] Allocca, C; d'Aquin, M; Motta, E. Detecting Different Versions of Ontologies in Large Ontology Repositories. 8_ Internacional Conferencia de Web Semántica, Chantilly, VA. USA. Páginas: 1- 14. 2009.