

Revisión Sistemática sobre Calidad de Datos en Portales Corporativos

M. Angélica Caro Coral Calero, Ismael Caballero, Mario Piattini

Dept. de Auditoría e Informática

Universidad del Bío Bío

La Castilla s/n

Chillán, Chile

mcaro@ubiobio.cl

Dept. de Informática

Universidad Castilla-La Mancha

Paseo de la Universidad 4

Ciudad Real, España

{Coral.Calero, Ismael.Caballero, Mario.Piattini}@uclm.es

Resumen

El avance tecnológico e Internet han propiciado la aparición de una gran diversidad de Aplicaciones Web (AW), entre ellas los Portales Corporativos (PC). A través de ellos las organizaciones desarrollan sus negocios en un entorno cada vez más competitivo. Un factor decisivo para esta competitividad es el uso de datos de calidad. Estos últimos años se han realizado diversas investigaciones en torno a la Calidad de los Datos (DQ), pero no existen muchos trabajos en el contexto de los PC. La revisión del estado del arte es esencial a la hora de comenzar cualquier proyecto de investigación y su desarrollo mediante una revisión sistemática permite evaluar e interpretar gran parte de la investigación relevante respecto de un tema en particular. En este artículo presentamos una adaptación de la propuesta de Kitchenham para revisiones sistemáticas, y publicamos los primeros resultados de nuestra revisión acerca de la DQ en PC.

1 Introducción

En los últimos 10 años se ha generado un creciente interés en el tema de DQ debido al incremento de la interconectividad de los productores y los consumidores de datos, estimulando principalmente por el desarrollo de Internet [18]. Una AW de gran evolución y mejor posicionamiento en estos últimos años son los Portales Web [11, 10]. El concepto de portal ha ido evolucionando hasta dar paso a complejas aplicaciones, entre ellas los Portales Corporativos.

Los PC permiten gestionar, en base a perfiles de usuario, el acceso a distintas aplicaciones y múltiples fuentes de información [33], simplifican la información compleja, proveen servi-

cios de aplicaciones útiles y en un contexto específico, fomentan la construcción de comunidades y colaboración a lo largo de toda la empresa u organización [11]. A través de los PC los usuarios pueden acceder a diversos tipos de AW, donde la DQ puede ser muy heterogénea ya que hay un conflicto entre la necesidad de publicar información de alta calidad y publicar la información tan pronto como esté disponible [35].

Nuestro interés de investigación está centrado en la DQ en los PC. Para esto es esencial partir por una revisión del estado del arte. Para darle mayor rigurosidad y pertinencia a nuestro trabajo, optamos por realizarla mediante una revisión sistemática de los trabajos en el área basada en [25]. Puesto que esa propuesta se orienta al trabajo de un equipo de investigadores, debimos adaptarla al trabajo de un solo investigador, en el contexto del desarrollo de una tesis doctoral. El objetivo de este artículo es dar a conocer dicha adaptación y los primeros resultados de nuestra revisión.

El resto del artículo se organiza de la siguiente forma: la sección 2 describe la metodología utilizada para el desarrollo de la revisión sistemática. La sección 3 corresponde a una síntesis del estado del arte. Y en la sección 4 exponemos nuestras conclusiones.

2 Metodología aplicada a la revisión de la literatura

La revisión del estado del arte es esencial para comenzar un proyecto de investigación. Nosotros optamos por realizarla mediante una revisión sistemática. Utilizamos como método una adaptación de la propuesta de Kitchenham [25], que pretende ser una guía para el desarro-

Tabla 1: Metodología para una revisión sistemática con un solo investigador.

Etapas	Planificación de la Revisión
	Identificar la necesidad de revisión Definir un protocolo de búsqueda Definir un protocolo de revisión Evaluar la planificación
Etapas	Desarrollo de la revisión
	Búsqueda de estudios primarios Selección de los estudios primarios Extracción y gestión de datos Síntesis de datos
Etapas	Publicación de los resultados

llo de revisiones sistemáticas dentro del área de la IS. Nosotros la adaptamos al trabajo de un solo investigador, en el contexto del desarrollo de una tesis doctoral que aborda la DQ en PC (tabla 1). Las siguientes subsecciones detallan cada etapa y explican cómo las abordamos.

2.1 Planificación de la revisión

Identificar la necesidad de la revisión.

Consiste en justificar la necesidad de realizar la revisión. Determinar aquí cuales son los objetivos de la revisión y con qué recursos se cuenta para comenzar (durante el desarrollo de ésta pudieran aparecer nuevos recursos).

De acuerdo con [25], las razones para emprender una revisión sistemática son: Resumir la evidencia existente concerniente a una tecnología, Identificar algún vacío en la investigación actual con el objeto de sugerir áreas para investigaciones futuras y Proveer un marco de trabajo y/o los antecedentes necesarios con el objeto de posicionar nuevas actividades de investigación.

En nuestro caso, el objetivo es resumir la evidencia existente concerniente a cómo se ha abordado la DQ en los PC o aplicaciones afines, para proveer un marco de trabajo y los antecedentes necesarios para desarrollar nuestro proyecto de investigación.

Las principales interrogantes que tratamos de resolver con nuestra revisión sistemática son: ¿Cuáles son los problemas asociados a la DQ en los PC (o en las AW)?, ¿Cuáles son las dimensiones de calidad asociadas a los datos en la Web?, ¿Cómo se puede medir la DQ en la Web?, ¿Cómo se puede mejorar la DQ en la Web?, ¿Qué herramientas existen para evaluar y/o mejorar la DQ en la Web?

Los recursos disponibles para acceder a los trabajos existentes en esta materia son:

- CD's con las actas de la Conferencia Internacional en Calidad de Datos (ICIQ) desde los años 1996 hasta 2004.
- Revistas electrónicas disponibles en la biblioteca virtual de la Universidad Castilla-La Mancha (UCLM).
- Libros de la biblioteca de UCLM, incluidos los en formato electrónico.
- Librería Digital de Literatura Científica (Citeseer), en <http://citeseer.ist.psu.edu/>
- Trabajos de investigación y Tesis doctorales relacionadas con nuestro tema.
- Textos y revistas disponibles en el laboratorio del grupo de investigación Alarcos de la UCLM.
- Internet, a través del uso de motores de búsqueda.

Definir un protocolo de búsqueda.

Consiste en definir las normas que seguirá la investigación respecto del proceso de búsqueda en las fuentes de información. Deben ser definidos los términos que se buscarán, las combinaciones de éstos, la estrategia de búsqueda empleada según cada fuente y cómo se registrarán los resultados. Es recomendable el registro de los resultados de las búsquedas ya que pueden servir posteriormente.

Este protocolo puede y debe ser mejorado en la medida que se avanza con la revisión. La tabla 2 muestra el protocolo de búsqueda definido en nuestra investigación.

Definir un protocolo de revisión.

Un protocolo de revisión especifica los métodos que serán usados para emprender una revisión sistemática que permita evitar prejuicios, de forma que la selección de los estudios no esté guiada por las expectativas del investigador [25]. Se deben definir las normas de revisión a seguir, los criterios de inclusión y exclusión que se emplearán, la estrategia de extracción de datos y finalmente la estrategia de síntesis.

En la tabla 3 se muestra el protocolo de revisión que nosotros empleamos. Considerando

Tabla 2: Protocolo de búsqueda.

Términos	Data, Information, Web, www, Internet, Quality, Portal, Corporate	
Combinaciones	"Data Quality" "Information Quality" "Data Quality"+ (Web/Portal/ "Corporate Portal"/Internet/www) "Information Quality"+ (Web/Portal /"Corporate Portal "/Internet/www)	
Estrategias de búsqueda	Recursos	Se pueden ingresar en forma escalada los términos de búsqueda, restringiéndola a los resultados de una búsqueda anterior.
	CD's de actas de congresos	Revisar los índices donde aparecen los títulos de cada artículo y el resumen, los cuales deberán ser leídos en busca de los términos y sus combinaciones.
	Texto electrónico	Una revisión manual de: título, resumen, índice de términos, etc.
	Internet	Cuando los resultados de una búsqueda nos llevan a sitios donde el documento que deseamos obtener está restringido, podemos buscar en las páginas personales de los autores.
	Autores	Acceder directamente a sus páginas personales o bien realizar búsqueda por sus nombres combinando con los términos definidos o bien dejar la combinación para un filtrado posterior.
	Artículos	Revisar las referencias bibliográficas que incluyen. En estos casos, se procederá a buscar directamente el documento citado usando los antecedentes que aparecen en la referencia.
	Otros	Se deberá considerar la posibilidad de la aparición de nuevos términos o conceptos que nos ayuden a encontrar trabajos de nuestro interés.
Registro de resultados	Los resultados de las búsquedas y el origen de los mismos, deben ser registrados por escrito.	

que la estructura básica de un estudio en formato de artículo científico es: Resumen, Introducción, Trabajos relacionados, Preliminares, Cuerpo del Artículo, Conclusión y Referencias.

Evaluación de la planificación.

Dado que esta investigación se enmarca en el desarrollo de una tesis doctoral, la evaluación fue realizada por los directores de la misma.

2.2 Desarrollo de la revisión

Durante esta etapa se lleva a cabo la revisión de los estudios primarios (estudios individuales que contribuyen a una revisión sistemática [25]) relacionados con el área de interés. Contempla 4 subetapas.

Búsqueda de estudios primarios

Se deberá realizar en base al protocolo de búsqueda definido (nosotros usamos el protocolo definido en la tabla 2).

Selección de estudios primarios

Esta debe hacerse en base al protocolo de revisión definido (en nuestro caso la tabla 3). Se puede registrar información acerca del proceso de selección, que pueda ser útil más adelante.

Extracción y gestión de datos

En esta etapa se extrae la información de interés en los estudios. Además debe registrarse la información necesaria para gestión, como la relativa a la bibliografía y otra que los investigadores consideren pertinente. Se realizará en base al protocolo de revisión definido (en nuestro caso la tabla 3 y la herramienta EndNote para mantener los datos más relevantes de cada estudio).

Síntesis de datos

Finalmente, la subetapa de síntesis de los datos consiste en aplicar la estrategia de síntesis definida en el protocolo de revisión. En nuestro caso implicó el abordar el estado del arte considerando los puntos especificados en la tabla 3. Por su extensión, los resultados de esta etapa se muestran en la sección 3.

2.3 Publicación de los resultados

Corresponde a la utilización de los resultados una vez que disponemos de ellos. Es importante comunicar los resultados de una revisión sistemática de manera efectiva, siendo la forma más común la publicación de un artículo, informe técnico o sección de una tesis [25].

Tabla 3: Protocolo de revisión de trabajos.

Normas de Revisión	Si no se dispone del trabajo completo este es el momento para conseguirlo, para asegurarnos de su relevancia para nuestra revisión. Se leerá al menos el resumen e introducción y se decidirá si incluimos o no el estudio, en base a los criterios de este protocolo. Se llevará un control acerca de los estudios primarios aceptados y rechazados, registrando el motivo de exclusión.
Criterios de Inclusión	Se incluirán todos aquellos trabajos o estudios que aborden el tema de DQ en PC y/o en la Web en algunos de los siguientes tópicos: - Análisis del problema de DQ en PC. - Propuesta de un marco de trabajo - Propuesta para evaluación/mejora - Herramientas
Criterios de Exclusión	Se excluirán aquellos estudios que a pesar de contener los términos de búsqueda o combinación de ellos, no contienen información relevante sobre el tema.
Estrategia de extracción de datos	Por cada estudio seleccionado, se extraerán datos en 2 etapas: 1: En esta podremos obtener la siguiente información : - A qué comunidad está orientado el artículo. [Introducción, Trabajos relacionados, Referencias] - Cuáles son sus contribuciones (según los autores). [Resumen, introducción, Conclusión] - Cuáles son las posibles consecuencias de las contribuciones (Aplicaciones directas, nuevas técnicas, nuevas áreas de investigación, etc.).[Introducción] 2: Se leerán los preliminares y el cuerpo del artículo, esto nos permitirá: - Incluir en forma detallada la información que necesitemos para nuestra revisión. - Comprender un experimento, los fundamentos de un marco de trabajo, las características de un modelo, etc.
Estrategia de síntesis de datos	Los datos serán sintetizados en: - Problemas de DQ en PC (o AW) - Propuestas de marcos de trabajo - Propuestas para evaluación/mejora - Herramientas

Tabla 4: Categorías y Dimensiones de DQ.

Categorías	Dimensiones
Intrínseca	Exactitud, Objetividad, Credibilidad, Reputación
Accesibilidad	Accesibilidad y Seguridad
Contextual	Relevancia, Valor agregado, Oportunidad, Completitud, Cantidad de Información
Representacional	Interpretabilidad, Fácil entendimiento, Representación Concisa, Representación consistente

En nuestro caso los resultados serán utilizados, primeramente, para conocer el estado del arte de nuestra área de interés y definir específicamente la investigación que abordaremos, como base para la preparación de artículos para revistas del área y/o congresos relacionados con el tema (como el presente) y para la realización de uno de los capítulos de la tesis doctoral en desarrollo.

3 Desarrollo de la Revisión Sistemática: Síntesis de Datos

En esta sección expondremos una parte de los resultados de la revisión sistemática. Nuestro interés es la DQ en los PC, sin embargo los resultados de nuestra revisión no nos han entregado muchos trabajos que desarrollen este tema, por lo que se optó por revisar los trabajos más relacionados, es decir DQ en la Web.

Un primer punto a considerar es que los términos "dato" e "información", son usados a menudo como sinónimos [40], y al realizar la revisión de la literatura relacionada nos encontramos con que se referencia indistintamente como "Calidad de Datos"(DQ) o "Calidad de Información"(IQ). Por el momento no haremos distinción entre dichos términos.

Por otro lado, en los últimos años se ha discutido mucho respecto de la relevancia de la Calidad de Datos (DQ/IQ) [41, 37, 40, 36] y en el contexto de los Sistemas de Información (SI) se han realizado diversas propuestas tendientes a definirla, evaluarla y mejorarla [40, 36, 6]. Por ejemplo uno de los marcos de trabajo más utilizados y/o referenciado es el propuesto por [37] que presenta una clasificación basada en la perspectiva del usuario, ver tabla 4.

En base a estas categorías y dimensiones, se han propuesto metodologías que permiten ges-

tionar la DQ en una organización, por ejemplo: TDQM [40], AIMQ [27] y TQdM [12].

Las AW representan para las organizaciones un vehículo para desarrollar diversos tipos de transacciones internas y externas, por lo cual no es sorprendente que exista un gran interés en evaluar la calidad de las AW y la DQ en ellas [26]. Dadas las características que las AW presentan, y su diferencia con los SI tradicionales, la comunidad de investigadores ha comenzado a tratar con gran interés el tema de DQ en la Web [18]. En las siguientes subsecciones sintetizaremos los resultados de nuestra revisión sistemática de acuerdo a como indica el protocolo de la tabla 3.

3.1 Problemas de DQ en la Web

Como afirman Huang et. al [22], la calidad de la información y del conocimiento resultan imprescindibles para el funcionamiento más competitivo de las organizaciones, sin embargo, el emplear Internet como portal de bienvenida dificulta su gestión. En nuestra revisión hemos identificado una serie de problemas asociados a la DQ en el contexto de la Web. Si bien éstos no están referidos específicamente a un PC podemos observar que dadas las características y funcionalidades que deben soportar éstos, son extensibles también a este tipo de AW: Problemas típicos de una página Web [14], Desarrollo del comercio electrónico [29, 11, 38, 20], Integración de datos estructurados y no estructurados [15], Integración de datos de fuentes diversas [32, 42, 18, 3, 5], DQ y la Perspectiva del Usuario [34, 3, 9, 18], Necesidad de entender los datos y su calidad [15, 18], AW y Web Semántica [4, 38, 39], Demanda de servicios en tiempo real [2], DQ Dinámica [35, 18], Fidelidad de los Usuarios [17].

3.2 Marcos de Trabajo

A través de un marco de trabajo de DQ las organizaciones pueden definir un modelo para el ambiente de datos, identificar atributos de DQ relevantes, analizar los atributos de DQ en el contexto actual y futuro, y proveer una guía para mejorar la DQ. En la literatura hemos encontrado algunos de ellos que están orientados a la DQ en la Web.

Naumann y Rolker [32], abordan la IQ desde 3 aspectos: el usuario, la fuente de datos y el proceso de consulta. Cada uno de estos se

desarrolla como fuente de metadatos de IQ, y a cada una de estas fuentes se le asocia una clase de métodos de evaluación.

Aboelmaged [1], plantea la necesidad de entender la IQ en el dominio del comercio electrónico y proporciona un modelo para abordarla, sin embargo, no plantea dimensiones ni criterios de IQ concretos.

Katerattanakul y Siau [23], adaptan las categorías de DQ propuestas en [37] al contexto Web. Luego en [24] proponen otro marco para identificar las características y facilidades de sitios Web de comercio electrónico en la Internet, enfocadas desde la perspectiva de los requerimientos de los usuarios Web.

Pernici y Scannapieco [35], definen una metodología y herramientas para la gestión de DQ en la construcción de AW. Consideraron cuatro dimensiones de DQ centradas en la naturaleza dinámica de los datos: expiración, completitud, confiabilidad de la fuente y exactitud.

Fugini et al. [16], abordan las AW como SI Cooperativos extendidos para proveer servicio electrónico en línea y proponen las dimensiones de calidad: Exactitud sintáctica y semántica, Completitud, Actualidad y Consistencia Interna. Además proponen dimensiones con las cuales la organización receptora es capaz de evaluar la DQ global de los datos recibidos: Oportunidad, Importancia, Fiabilidad de la fuente y Confidencialidad.

En [19] Graefe destaca la importancia de la IQ en Internet en el contexto de la toma de decisiones, y muestra cómo la IQ es consecuencia de la conducta de quienes la proveen y reciben. Proponen ocho criterios de IQ asociados al Proceso de Decisión: Disponibilidad, Interpretabilidad, Novedad, Credibilidad, Relevancia, Disponibilidad, Pre-decisión, Validez, Valor de la Información.

Eppler [13] propone un marco para evaluar la calidad de un sitio Web, y la satisfacción del usuario con respecto a éste. Contempla 16 atributos de IQ, agrupados en 4 dimensiones: Relevancia o comunidad, Intrínseca o de legitimidad, Proceso e Infraestructura. El modelo identifica responsables de la IQ: administradores, autores o proveedores de información, Webmaster y administradores de la tecnología de información.

En [31] Moustakis considera, para la eva-

luación de sitios Web, los criterios de calidad: Relevancia, Utilidad, Fiabilidad, Especialización, Arquitectura, Navegabilidad, Eficiencia, Diseño y Animación. Los 4 primeros directamente relacionados con la DQ.

Melkas [30] propone un marco para el análisis de la IQ en procesos de información de redes organizacionales complejas en base a datos cualitativos. Está estructurado en 6 etapas de análisis: Bases, Componentes, Contenido y recursos, Oportunidad, Enrutamiento y Procesamiento de Procedimientos. Cada estado de análisis representa un criterio de IQ clave para la operación exitosa de una red.

Bouzeghoub y Peralta [5], abordan la problemática de la evaluación de la frescura de los datos (asociada a los factores: Actualidad y Oportunidad) en el contexto de los sistemas de integración de datos, los cuales integran datos de fuentes diferentes y proveen a los usuarios acceso uniforme a los datos por medio de un modelo global.

Las propuestas señaladas, no abordan la DQ explícitamente en PC, pero los aspectos que ellas mencionan son totalmente válidos en este contexto aunque podrían no ser suficientes para representar la DQ en los PC.

3.3 Métodos de Evaluación

Una vez que se dispone de un marco de trabajo, pueden desarrollarse métodos de evaluación y mejora de la DQ. Muchos de los siguientes trabajos se basan en los marcos anteriores.

En [23] proponen un instrumento de evaluación de IQ para sitios Web individuales.

Zhu y Buchmann proponen un conjunto de criterios para la evaluación y selección de recursos Web como fuentes de datos externas de un Almacén de Datos [42].

En [14] se distinguen los tipos de herramientas que pueden ser usadas para medir DQ en la Web y proponen la metodología IQM para aplicarlas correctamente.

En el contexto de los sistemas de información financieros de arquitectura multi-canal, en [7] elaboran un método de medición del criterio de DQ *Actualidad*. Los mismos investigadores en [8] proponen una metodología para soportar la selección de la información relevante para ser preservada y las técnicas de evaluación de DQ son aplicadas para definir y medir la relevancia de la información Web. Luego en [9], plantean

un modelo que une la fase de evaluación a los requerimientos de calidad de los usuarios.

En [28] se describe un componente de la arquitectura DaQuinCIS: Servicio de Notificación de Calidad (QNS), usado para informar a los usuarios interesados cuando cambian los valores de calidad en un SI cooperativo.

En [21] se presenta una metodología cualitativa y orientada al usuario basada en computación difusa que genera recomendaciones lingüísticas de la calidad informativa de sitios Web, basada en las percepciones de usuarios.

En [3] se propone DQM un administrador de DQ para establecer comunicación entre los procesos de integración de información, los usuarios y la aplicación; para manejar la heterogeneidad semántica y la DQ.

Los métodos propuestos son muy diversos y no están orientados a PC. Además como se plantea en [9] la literatura no provee un conjunto exhaustivo de métricas que las organizaciones puedan aplicar.

3.4 Conclusiones de la revisión

La gran mayoría de los trabajos en la literatura muestran que la DQ es muy dependiente del contexto. El aumento del interés por el desarrollo de aplicaciones en la Web ha implicado nuevas propuestas de marcos de trabajo, metodologías y métodos de evaluación de DQ o bien la adaptación de los ya existentes en otros contextos como por ejemplo: SI tradicionales, SI Cooperativos y Almacenes de Datos.

En esta sección hemos descrito diversos problemas asociados a la DQ en la Web, muchos de los cuales aún despiertan gran interés por parte de la comunidad de investigadores. Existen diversas propuestas en relación a la DQ en la Web.

Como conclusión de nuestra revisión bibliográfica creemos que la DQ en el contexto de los Portales Web, y en particular en PC, no ha sido abordada en forma específica. Consideramos pertinente y necesario el desarrollo de trabajos específicos a este contexto y, consecuentemente con ello, nuestra propuesta de investigación se orienta al desarrollo de un *Modelo de Calidad de Datos para Portales Corporativos*. Considerando que un portal es un Sistema de Integración de Aplicaciones Web Heterogéneas, creemos que los trabajos existentes en la literatura relacionada constituirán una buena base

para nuestra investigación.

4 Conclusiones

El desarrollo del estado del arte es una etapa fundamental en un trabajo de investigación y la realización de una revisión sistemática provee mayor rigurosidad a ésta. La definición de una metodología para el desarrollo de la revisión sistemática y su posterior aplicación a nuestro trabajo nos ha ayudado a organizar de mejor forma nuestro trabajo actual y futuro. también nos ha permitido obtener valiosa información de interés para nuestra investigación acerca de la DQ en PC y poder:

- Resumir la evidencia encontrada.
- Identificar vacíos en la investigación actual para sugerir áreas para investigaciones futuras.
- Proveer un marco de trabajo y/o los antecedentes necesarios con el objeto de posicionar nuevas actividades de investigación.

Nuestros próximos pasos son completar el estado del arte expuesto en este artículo y sentar las bases de nuestra propuesta de investigación, en el contexto del desarrollo de una tesis doctoral en torno al desarrollo de un Modelo de Calidad de Datos para Portales Corporativos.

5 Agradecimientos

Esta investigación es parte del proyecto CALIPO (TIC 2003-07804-C05-03) del Ministerio de Ciencia y Tecnología y del proyecto MESSENGER (PCC-03-003-1) de la Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha.

Referencias

- [1] M. Aboelmegeed. A soft system perspective on information quality in electronic commerce. In *Proceeding of ICIQ-2000*, pages 318–319, 2000.
- [2] M. Amirijoo, J. Hansson, and S. Son. Specification and management of QoS in imprecise real-time databases. In IEEE, editor, *Proceeding of the Seventh International Database Engineering and Applications Symposium*, pages 192–201, 2003.
- [3] P. Angeles and L. MacKinnon. Detection and resolution of data inconsistencies, and data integration using data quality criteria. In *QUATIC'2004*, pages 87–93, 2004.
- [4] T. Bernes-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 2001.
- [5] M. Bouzeghoub and V. Peralta. A framework for analysis of data freshness. In *Proceeding of IQIS2004*, pages 59–67, Paris, France, 2004. ACM.
- [6] I. Caballero, M. Piattini, and M. A. Blanco. Mejora en la gestión de calidad de la información. *GTI*, 2004.
- [7] C. Cappiello, C. Francalanci, and B. Pernici. A model of data currency in multi-channel financial architectures. In *Proceeding of ICIQ 2002*, pages 106–118, 2002.
- [8] C. Cappiello, C. Francalanci, and B. Pernici. Preserving web sites: A data quality approach. In *Proceeding of ICIQ 2003*, pages 331–343, 2003.
- [9] C. Cappiello, C. Francalanci, and B. Pernici. Data quality assessment from the user's perspective. In *Proceeding of IQIS2004*, pages 68–73, Paris, Francia, 2004. ACM.
- [10] H. Collins. *Corporate Portal Definition and Features*. AMACOM, 2001.
- [11] M. Davydov. *Corporate Portals and e-Business Integration*. Emerging Business Technology Series. McGraw-Hill, 2001.
- [12] L. English. Total quality data management (TQdM). In M. Piattini, C. Calero, and M. Genero, editors, *Information and Database Quality*. Kluwer Academic Publishers, 2001.
- [13] M. Eppler, R. Algesheimer, and M. Dimpfel. Quality criteria of content-driven websites and their influence on customer satisfaction and loyalty: An empirical test of an information quality framework. In *Proceeding of ICIQ2003*, pages 108–120, 2003.
- [14] M. Eppler and P. Muenzenmayer. Measuring information quality in the web context: A survey of state-of-the-art instruments and an application methodology. In *Proceeding of ICIQ2002*, pages 187–196, 2002.
- [15] C. Finkelstein and P. Aiken. Xml and corporate portals, Revisado el 02/02/2005 1999.
- [16] M. Fugini, M. Mecella, P. Plebani, B. Pernici, and M. Scannapieco. Data quality in cooperative web information systems. 2002.
- [17] J. C. Garcia. Portales de internet: concepto, tipología básica y desarrollo. *El profesional de la información*, 10, nº 7-8:4–13, 2001.
- [18] M. Gertz, T. Ozsu, G. Saake, and K.-U. Sattler. Report on the dagstuhl seminar "data quality on the web". *SIGMOD Record*, vol. 33, Nº 1:127–132, 2004.
- [19] G. Graefe. Incredible information on the internet: Biased information provision and a lack of credibility as a cause of insufficient information quality. In *Proceeding of ICIQ2003*, pages 133–146, 2003.
- [20] A. Haider and A. Koronios. Authenticity of information in cyberspace: IQ in the internet, web, and e-business. In *Proceeding of ICIQ2003*, pages 121–132, 2003.

- [21] E. Herrera-Viedma, G. Pasi, and A. Lopez-Herrera. Evaluating the information quality of web sites: A quality methodology based on fuzzy computing with words. Technical report, 2004.
- [22] K.-T. Huang, Y. Lee, and R. Wang. *Calidad de la informacion y gestion del conocimiento*. AENOR, 2000.
- [23] P. Katerattanakul and K. Siau. Measuring information quality of web sites: Development of an instrument. In *Proceeding of the 20th International Conference on Information System*, pages 279–285, 1999.
- [24] P. Katerattanakul and K. Siau. Information quality in internet commerce desing. In M. Piattini, C. Calero, and M. Genero, editors, *Information and Database Quality*. Kluwer Academic Publishers, 2001.
- [25] B. Kitchenham. Procedures for performing systematic reviews. Technical Report 0400011T.1, 2004.
- [26] D. Pcopso, L. Pipino, and W. Rybolt. The assessment of web site quality. In *Proceeding of ICIQ2000*, pages 97–108, 2000.
- [27] Y. Lee. AIMQ: a methodology for information quality assessment. *Information and Management. Elsevier Science*, pages 133–146, 2002.
- [28] C. Marchetti, M. Mecella, M. Scannapieco, and A. Virgillito. Enabling data quality notification in cooperative information systems through a web-service based architecture. In *Proceeding of the Fourth International Conference on Web Information Systems Engineering*, pages 329–332, 2003.
- [29] J. McNamara. Cleaning up web data. *DB2 Magazine*, pages 14–24, Winter 1999 1999.
- [30] H. Melkas. Analyzing information quality in virtual service networks with qualitative interview data. In *Proceeding of ICIQ2004*, pages 74–88, 2004.
- [31] V. Moustakis, C. Litos, A. Dalivigas, and L. Tsiro-nis. Website quality assessment criteria. In *Proceeding of ICIQ2004*, pages 59–73, 2004.
- [32] F. Naumann and C. Rolker. Assessment methods for information quality criteria. In *Proceeding of ICIQ2000*, pages 148–162, 2000.
- [33] Y. Pantoja, Y. Cervantes, and F. Perez. Los portales corporativos. *Telem@tica*, Año 3, N° 16(1729-3804):3–6, 2004.
- [34] R. Pautke and T. Redman. The organization's most important data issues. In M. Piattini, C. Calero, and M. Genero, editors, *Information and Database Quality*. Kluwer Academic Publishers, 2001.
- [35] B. Pernici and M. Scannapieco. Data quality in web information systems. In *Proceeding of the 21st International Conference on Conceptual Modeling*, pages 397–413, 2002.
- [36] L. Pipino, Y. Lee, and R. Wang. Data quality assessment. *Communications of the ACM*, Vol. 45, N° 4:211–218, 2002.
- [37] D. Strong, Y. Lee, and R. Wang. Data quality in context. *Communications of the ACM*, Vol. 40, N° 5:103–110, 1997.
- [38] B. Thuraisingham. Building secure survivable semantic webs. In *Proceeding of the 14th IEEE International Conference on Tools with Artificial Intelligence*, pages 395–398, 2002-a.
- [39] B. Thuraisingham, E. Hughes, and D. Allen. Dependable semantic web. In *IEEE WORDS 2002*, San Diego, CA, 2002-b.
- [40] R. Wang. A product perspective on total data quality management. *Communications of the ACM*, Vol. 41, N° 2:54–65, 1998.
- [41] R. Wang, V. Storey, and C. Firth. A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7, N° 4, 1995.
- [42] Y. Zhu and A. Buchmann. Evaluating and selecting web sources as external information resources of a data warehouse. In *Proceeding of the 3rd International Conference on Web Information Systems Engineering*, pages 149–160, 2002.