

Chapter 3.19

A Data Quality Model for Web Portals

Angélica Caro

University of Bio Bio, Chile

Coral Calero

University of Castilla-La Mancha, Spain

Mario Piattini

University of Castilla-La Mancha, Spain

ABSTRACT

Web portals are Internet-based applications that provide a big amount of data. The data consumer who uses the data given by these applications needs to assess data quality. Due to the relevance of data quality on the Web together with the fact that DQ needs to be assessed within the context in which data are generated, data quality models specific to this context are necessary. In this chapter, we will introduce a model for data quality in Web portals (PDQM). PDQM has been built upon the foundation of three key aspects: (1) a set of Web data quality attributes identified in the literature in this area, (2) data quality expectations of data consumers on the Internet, and (3) the functionalities that a Web portal may offer its users.

DOI: 10.4018/978-1-59904-847-5.ch008

INTRODUCTION

In recent years, Web portals have risen in popularity as a way of aggregating, organizing, and presenting content in a highly uniform, customizable, and personalized way. In simplest terms, a portal is a Web site that provides content and application functionality in a way that is both useful and meaningful to the end user (Secrist 2003).

In general, Web portals provide users with access to different data sources (providers) (Mahdavi, Shepherd, Benatallah, 2004), as well as to online information and information-related services (Yang, Cai, Zhou, & Zhou, 2004). Moreover, they create a working environment where users can easily navigate in order to find the information they specifically need to quickly perform their operational or strategic tasks and make decisions (Collins, 2001). So, users

or data consumers aimed at using the data offered by these applications need to ensure that these data are appropriate for the use they need, being fundamental to assess the quality of data.

Data and information quality (DQ hereafter) is often defined as “fitness for use,” that is, the ability of a data collection to meet user requirements (Strong, Lee, & Wang 1997; Cappiello, Francalanci, & Pernici, 2004). This definition suggests the relativity of this concept because data considered appropriate for a specific use may not be appropriate for another. Even more, this definition involves understanding DQ from the user’s point of view and, consequently, understanding that the quality of data cannot be assessed independently of the people who use data (Wang & Strong, 1996).

Due to the relevance of DQ on the Web together with the fact that DQ needs to be assessed within the context of its generations (Knight & Burn, 2005), in the last years the research community started studying the subject of DQ in the Web context (Gertz, Ozsu, Sattke, & Sattler, 2004). However, despite the sizeable body of literature available on DQ and the different domains studied on the Web, we have found no works on DQ that address the particular context of Web portals. Likewise, except for a few works in the DQ area, such as (Wang & Strong, 1996; Burgess, Fiddian, & Gray, 2004; Cappiello et al., 2004), most of the contributions target quality from the data producers’ or data custodians’ perspective and not from the data consumers’ perspective (Burgess, et al., 2004). The last perspective differs from the two others in two important aspects: (1) data consumers have no control over the quality of available data and (2) the aim of consumers is to find data that match their personal needs, rather than provide data that meet the needs of others.

So, consequently to this situation, the aim of our research work is the creation of a Data Quality Model for Web Portals (PDQM). The objective of this chapter is to present the definition of PDQM. This model is focused on the data consumer’s perspective and as key pieces in its

definition we have taken: (1) a set of Web DQ attributes identified in the literature, (2) the DQ expectations of data consumers on the Internet described by Redman (Redman, 2000) and (3) the functionalities that a portal Web may offer its users (Collins, 2001).

BACKGROUND

Data Quality and the Web

Research on DQ began in the context of information systems (Strong et al., 1997; Lee, 2002) and it has been extended to contexts such as cooperative systems, data warehouses or e-commerce, among others. Due to the particular characteristics of Web applications and their differences from the traditional information systems, the research community started to deal with the subject of DQ on the Web (Gertz et al., 2004). In fact, the particular nature of the Internet has forced to pay attention to a series of particular issues of this context that can affect or influence the quality of data. We have summarized some of them in Table 1.

In the last years, based on the previous Web issues and others, frameworks and models to deal with DQ in different domains in the Web context have been proposed. Among them we can highlight those shown in Table 2.

Concerning Table 2, we can make two important observations. First, the frameworks proposed tackle different domains on the Web. This reasserts the idea that DQ needs to be assessed within the context of the data source (Knight & Burn, 2005). Second, for Web portals we have not found specific DQ frameworks.

During the past decade, an increasing number of organizations have established Web portals to complement, substitute, or widen existing services to their clients. In general, portals provide users with access to different data sources (providers) (Mahdavi, et al., 2004), as well as to online information and information-related services (Yang et al., 2004).

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/data-quality-model-web-portals/43984

Related Content

QoS-Aware Web Services Discovery with Federated Support for UDDI

Chen Zhou, Liang-Tien Chiaand Bu-Sung Lee (2010). *Electronic Services: Concepts, Methodologies, Tools and Applications* (pp. 605-627).

www.irma-international.org/chapter/qos-aware-web-services-discovery/43973

Cloud Computing: The Future of Big Data Management

Shimaa Oufand Mona Nasr (2015). *International Journal of Cloud Applications and Computing* (pp. 53-61).

www.irma-international.org/article/cloud-computing/127105

Software Implementation of Real-time Discrete Wavelet Transform Algorithm with Filter Banks

Nikolajs Bogdanovs, Elans Grabsand Ernests Petersons (2016). *International Journal of Information Systems in the Service Sector* (pp. 70-86).

www.irma-international.org/article/software-implementation-of-real-time-discrete-wavelet-transform-algorithm-with-filter-banks/149189

Database Sharding: To Provide Fault Tolerance and Scalability of Big Data on the Cloud

Sikha Baguiand Loi Tang Nguyen (2015). *International Journal of Cloud Applications and Computing* (pp. 36-52).

www.irma-international.org/article/database-sharding/127104

Network Strategies of Hospitality Companies in Emerging and Transitory Economies: Evidence from Russia

Olga Balaeva, Ella Burnatseva, Marina Predvoditeleva, Marina Shereshevaand Olga Tretyak (2012). *Service Science Research, Strategy and Innovation: Dynamic Knowledge Management Methods* (pp. 519-546).

www.irma-international.org/chapter/network-strategies-hospitality-companies-emerging/61893