Lecture Notes in Computer Science

The LNCS series reports state-of-the-art results in computer science research, development, and education, at a high level and in both printed and electronic form. Enjoying tight cooperation with the R&D community, with numerous individuals, as well as with prestigious organizations and societies, LNCS has grown into the most comprehensive computer science research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology including interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

More recently, several color-cover sublines have been added featuring, beyond a collection of papers, various added-value components; these sublines include

- tutorials (textbook-like monographs or collections of lectures given at advanced courses)
- state-of-the-art surveys (offering complete and mediated coverage of a topic)
- hot topics (introducing emergent topics to the broader community)

In parallel to the printed book, each new volume is published electronically in LNCS Online.

Detailed information on LNCS can be found at
www.springer.com/lncs

Proposals for publication should be sent to
LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany
E-mail: lncs@springer.com

ISSN 0302-9743

Lecture Notes in Computer Science
LNCS LNAI LNBI

springer.com

Gavrilova et al. (Eds.)

LNCS 3982

ICCSA 2006

Marina Gavrilova et al. (Eds.)

Computational Science and Its Applications – ICCSA 2006

International Conference
Glasgow, UK, May 2006
Proceedings, Part III

3 Part III

Springer

Lecture Notes in Computer Science 3982

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Marina Gavrilova Osvaldo Gervasi
Vipin Kumar C.J. Kenneth Tan
David Taniar Antonio Laganà
Youngsong Mun Hyunseung Choo (Eds.)

Computational Science and Its Applications – ICCSA 2006

International Conference
Glasgow, UK, May 8-11, 2006
Proceedings, Part III

Springer

Volume Editors

Marina Gavrilova
University of Calgary, Canada
E-mail: marina@cpsc.ucalgary.ca

Osvaldo Gervasi
University of Perugia, Italy
E-mail: ogervasi@computer.org

Vipin Kumar
University of Minnesota, Minneapolis, USA
E-mail: kumar@cs.umn.edu

C.J. Kenneth Tan
OptimaNumerics Ltd., Belfast, UK
E-mail: cjtan@optimanumerics.com

David Taniar
Monash University, Clayton, Autralia
E-mail: david.taniar@infotech.monash.edu.au

Antonio Laganà
University of Perugia, Italy
E-mail: lag@unipg.it

Youngsong Mun
SoongSil University, Seoul, Korea
E-mail: mun@computing.soongsil.ac.kr

Hyunseung Choo
Sungkyunkwan University, Suwon, Korea
E-mail: choo@ece.skku.ac.kr

Library of Congress Control Number: 2006925086

CR Subject Classification (1998): F, D, G, H, I, J, C.2-3

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-540-34075-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-34075-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11751595 06/3142 5 4 3 2 1 0

Preface

This five-volume set was compiled following the 2006 International Conference on Computational Science and its Applications, ICCSA 2006, held in Glasgow, UK, during May 8–11, 2006. It represents the outstanding collection of almost 664 refereed papers selected from over 2,450 submissions to ICCSA 2006.

Computational science has firmly established itself as a vital part of many scientific investigations, affecting researchers and practitioners in areas ranging from applications such as aerospace and automotive, to emerging technologies such as bioinformatics and nanotechnologies, to core disciplines such as mathematics, physics, and chemistry. Due to the shear size of many challenges in computational science, the use of supercomputing, parallel processing, and sophisticated algorithms is inevitable and becomes a part of fundamental theoretical research as well as endeavors in emerging fields. Together, these far-reaching scientific areas contributed to shaping this conference in the realms of state-of-the-art computational science research and applications, encompassing the facilitating theoretical foundations and the innovative applications of such results in other areas.

The topics of the refereed papers span all the traditional as well as emerging computational science realms, and are structured according to the five major conference themes:

- Computational Methods, Algorithms and Applications
- High-Performance Technical Computing and Networks
- Advanced and Emerging Applications
- Geometric Modeling, Graphics and Visualization
- Information Systems and Information Technologies

Moreover, submissions from 31 workshops and technical sessions in areas such as information security, mobile communication, grid computing, modeling, optimization, computational geometry, virtual reality, symbolic computations, molecular structures, Web systems and intelligence, spatial analysis, bioinformatics and geocomputations, are included in this publication. The continuous support of computational science researchers has helped ICCSA to become a firmly established forum in the area of scientific computing.

We recognize the contribution of the International Steering Committee and sincerely thank the International Program Committee for their tremendous support in putting this conference together, the near 800 referees for their diligent work, and the IEE European Chapter for their generous assistance in hosting the event.

We also thank our sponsors for their continuous support without which this conference would not be possible.

Finally, we thank all authors for their submissions and all invited speakers and conference attendants for making the ICCSA Conference truly one of the premium events on the scientific community scene, facilitating exchange of ideas, fostering new collaborations, and shaping the future of computational science.

May 2006

Marina L. Gavrilova
Osvaldo Gervasi

on behalf of the co-editors
Vipin Kumar
Chih Jeng Kenneth Tan
David Taniar
Antonio Laganà
Youngsong Mun
Hyunseung Choo

Organization

ICCSA 2006 was organized by the Institute of Electrical Engineers (IEE)(UK), the University of Perugia (Italy), Calgary University (Canada) and Minnesota University (USA).

Conference Chairs

Vipin Kumar (University of Minnesota, Minneapolis, USA), Honorary Chair
Marina L. Gavrilova (University of Calgary, Calgary, Canada), Conference Co-chair, Scientific
Osvaldo Gervasi (University of Perugia, Perugia, Italy), Conference Co-chair, Program

Steering Committee

Vipin Kumar (University of Minnesota, USA)
Marina L. Gavrilova (University of Calgary, Canada)
Osvaldo Gervasi (University of Perugia, Perugia, Italy)
C. J. Kenneth Tan (OptimaNumerics, UK)
Alexander V. Bogdanov (Institute for High Performance Computing and Data Bases, Russia)
Hyunseung Choo (Sungkyunkwan University, Korea)
Andres Iglesias (University of Cantabria, Spain)
Antonio Laganà (University of Perugia, Italy)
Heow-Pueh Lee (Institute of High Performance Computing, Singapore)
Youngsong Mun (Soongsil University, Korea)
David Taniar (Monash University, Australia)

Workshop Organizers

Applied Cryptography and Information Security (ACIS 2006)

Sherman S.M. Chow (New York University, USA)
Joseph K. Liu (University of Bristol, UK)
Patrick Tsang (Dartmouth College, USA)
Duncan S Wong (City University of Hong Kong, Hong Kong)

Approaches or Methods of Security Engineering (AMSE 2006)

Haeng Kon Kim (Catholic University of Daegu, Korea)
Tai-hoon Kim (Korea Information Security Agency, Korea)

Authentication, Authorization and Accounting (AAA 2006)
Haeng Kon Kim (Catholic University of Daegu, Korea)

Computational Geometry and Applications (CGA 2006)
Marina Gavrilova (University of Calgary, Calgary, Canada)

Data Storage Devices and Systems (DSDS 2006)
Yeonseung Ryu (Myongji University, Korea)
Junho Shim (Sookmyong Womens University, Korea)
Youjip Won (Hanyang University, Korea)
Yongik Eom (Seongkyunkwan University, Korea)

Embedded System for Ubiquitous Computing (ESUC 2006)
Tei-Wei Kuo (National Taiwan University, Taiwan)
Jiman Hong (Kwangwoon University, Korea)

4th Technical Session on Computer Graphics (TSCG 2006)
Andres Iglesias (University of Cantabria, Spain)
Deok-Soo Kim (Hanyang University, Korea)

GeoComputation (GC 2006)
Yong Xue (London Metropolitan University, UK)

Image Processing and Computer Vision (IPCV 2006)
Jiawan Zhang (Tianjin University, China)

Intelligent Services and the Synchronization in Mobile Multimedia Networks (ISS 2006)
Dong Chun Lee (Howon University, Korea)
Kuinam J Kim (Kyonggi University, Korea)

Integrated Analysis and Intelligent Design Technology (IAIDT 2006)
Jae-Woo Lee (Konkuk University, Korea)

Information Systems Information Technologies (ISIT 2006)
Youngsong Mun (Soongsil University, Korea)

Information Engineering and Applications in Ubiquitous Computing Environments (IEAUCE 2006)
Sangkyun Kim (Yonsei University, Korea)
Hong Joo Lee (Dankook University, Korea)

Internet Communications Security (WICS 2006)
Sierra-Camara Josè Maria (University Carlos III of Madrid, Spain)

Mobile Communications (MC 2006)
Hyunseung Choo (Sungkyunkwan University, Korea)

Modelling Complex Systems (MCS 2006)
John Burns (Dublin University, Ireland)
Ruili Wang (Massey University, New Zealand)

Modelling of Location Management in Mobile Information Systems (MLM 2006)
Dong Chun Lee (Howon University, Korea)

Numerical Integration and Applications (NIA 2006)
Elise de Doncker (Western Michigan University, USA)

Specific Aspects of Computational Physics and Wavelet Analysis for Modelling Suddenly-Emerging Phenomena in Nonlinear Physics, and Nonlinear Applied Mathematics (PULSES 2006)
Carlo Cattani (University of Salerno, Italy)
Cristian Toma (Titu Maiorescu University, Romania)

Structures and Molecular Processes (SMP 2006)
Antonio Laganà (University of Perugia, Perugia, Italy)

Optimization: Theories and Applications (OTA 2006)
Dong-Ho Lee (Hanyang University, Korea)
Deok-Soo Kim (Hanyang University, Korea)
Ertugrul Karsak (Galatasaray University, Turkey)

Parallel and Distributed Computing (PDC 2006)
Jiawan Zhang (Tianjin University, China)

Pattern Recognition and Ubiquitous Computing (PRUC 2006)
Jinok Kim (Daegu Haany University, Korea)

Security Issues on Grid/Distributed Computing Systems (SIGDCS 2006)
Tai-Hoon Kim (Korea Information Security Agency, Korea)

Technologies and Techniques for Distributed Data Mining (TTDDM 2006)
Mark Baker (Portsmouth University, UK)
Bob Nichol (Portsmouth University, UK)

Ubiquitous Web Systems and Intelligence (UWSI 2006)
David Taniar (Monash University, Australia)
Eric Pardede (La Trobe University, Australia)

Ubiquitous Application and Security Service (UASS 2006)
Yeong-Deok Kim (Woosong University, Korea)

Visual Computing and Multimedia (VCM 2006)
Abel J. P. Gomes (University Beira Interior, Portugal)

Virtual Reality in Scientific Applications and Learning (VRSAL 2006)
Osvaldo Gervasi (University of Perugia, Italy)
Antonio Riganelli (University of Perugia, Italy)

Web-Based Learning (WBL 2006)
Woochun Jun Seoul (National University of Education, Korea)

Program Committee

Jemal Abawajy (Deakin University, Australia)
Kenny Adamson (EZ-DSP, UK)
Srinivas Aluru (Iowa State University, USA)
Mir Atiqullah (Saint Louis University, USA)
Frank Baetke (Hewlett Packard, USA)
Mark Baker (Portsmouth University, UK)
Young-Cheol Bang (Korea Polytechnic University, Korea)
David Bell (Queen's University of Belfast, UK)
Stefania Bertazzon (University of Calgary, Canada)
Sergei Bespamyatnikh (Duke University, USA)
J. A. Rod Blais (University of Calgary, Canada)
Alexander V. Bogdanov (Institute for High Performance Computing and Data Bases, Russia)
Peter Brezany (University of Vienna, Austria)
Herve Bronnimann (Polytechnic University, NY, USA)
John Brooke (University of Manchester, UK)
Martin Buecker (Aachen University, Germany)
Rajkumar Buyya (University of Melbourne, Australia)
Jose Sierra-Camara (University Carlos III of Madrid, Spain)
Shyi-Ming Chen (National Taiwan University of Science and Technology, Taiwan)
YoungSik Choi (University of Missouri, USA)
Hyunseung Choo (Sungkyunkwan University, Korea)
Bastien Chopard (University of Geneva, Switzerland)
Min Young Chung (Sungkyunkwan University, Korea)
Yiannis Cotronis (University of Athens, Greece)
Danny Crookes (Queen's University of Belfast, UK)
Jose C. Cunha (New University of Lisbon, Portugal)
Brian J. d'Auriol (University of Texas at El Paso, USA)
Alexander Degtyarev (Institute for High Performance Computing and Data Bases, Russia)
Frederic Desprez (INRIA, France)
Tom Dhaene (University of Antwerp, Belgium)
Beniamino Di Martino (Second University of Naples, Italy)
Hassan Diab (American University of Beirut, Lebanon)
Ivan Dimov (Bulgarian Academy of Sciences, Bulgaria)
Iain Duff (Rutherford Appleton Laboratory, UK and CERFACS, France)
Thom Dunning (NCSA and University of Illinois, USA)
Fabrizio Gagliardi (Microsoft, USA)
Marina L. Gavrilova (University of Calgary, Canada)
Michael Gerndt (Technical University of Munich, Germany)
Osvaldo Gervasi (University of Perugia, Italy)
Bob Gingold (Australian National University, Australia)
James Glimm (SUNY Stony Brook, USA)

Christopher Gold (Hong Kong Polytechnic University, Hong Kong)
Yuriy Gorbachev (Institute of High Performance Computing and Information Systems, Russia)
Andrzej Goscinski (Deakin University, Australia)
Jin Hai (Huazhong University of Science and Technology, China)
Ladislav Hluchy (Slovak Academy of Science, Slovakia)
Xiaohua Hu (Drexel University, USA)
Eui-Nam John Huh (Seoul Women's University, Korea)
Shen Hong (Japan Advanced Institute of Science and Technology, Japan)
Paul Hovland (Argonne National Laboratory, USA)
Andres Iglesias (University of Cantabria, Spain)
Peter K. Jimack (University of Leeds, UK)
In-Jae Jeong (Hanyang University, Korea)
Chris Johnson (University of Utah, USA)
Benjoe A. Juliano (California State University at Chico, USA)
Peter Kacsuk (MTA SZTAKI Researc Institute, Hungary)
Kyung Wo Kang (KAIST, Korea)
Carl Kesselman (USC/ Information Sciences Institute, USA)
Daniel Kidger (Quadrics , UK)
Haeng Kon Kim (Catholic University of Daegu, Korea)
Jin Suk Kim (KAIST, Korea)
Tai-Hoon Kim (Korea Information Security Agency, Korea)
Yoonhee Kim (Syracuse University, USA)
Mike Kirby (University of Utah, USA)
Dieter Kranzlmueller (Johannes Kepler University Linz, Austria)
Deok-Soo Kim (Hanyang University, Korea)
Vipin Kumar (University of Minnesota, USA)
Domenico Laforenza (Italian National Research Council, Italy)
Antonio Laganà (University of Perugia, Italy)
Joseph Landman (Scalable Informatics LLC, USA)
Francis Lau (The University of Hong Kong, Hong Kong)
Bong Hwan Lee (Texas A&M University, USA)
Dong Chun Lee (Howon University, Korea)
Dong-Ho Lee (Institute of High Performance Computing, Singapore)
Sang Yoon Lee (Georgia Institute of Technology, USA)
Tae-Jin Lee (Sungkyunkwan University, Korea)
Bogdan Lesyng (ICM Warszawa, Poland)
Zhongze Li (Chinese Academy of Sciences, China)
Laurence Liew (Scalable Systems Pte, Singapore)
David Lombard (Intel Corporation, USA)
Emilio Luque (University Autonoma of Barcelona, Spain)
Michael Mascagni (Florida State University, USA)
Graham Megson (University of Reading, UK)
John G. Michopoulos (US Naval Research Laboratory, USA)
Edward Moreno (Euripides Foundation of Marilia, Brazil)

Youngsong Mun (Soongsil University, Korea)
Jiri Nedoma (Academy of Sciences of the Czech Republic, Czech Republic)
Genri Norman (Russian Academy of Sciences, Russia)
Stephan Olariu (Old Dominion University, USA)
Salvatore Orlando (University of Venice, Italy)
Robert Panoff (Shodor Education Foundation, USA)
Marcin Paprzycki (Oklahoma State University, USA)
Gyung-Leen Park (University of Texas, USA)
Ron Perrott (Queen's University of Belfast, UK)
Dimitri Plemenos (University of Limoges, France)
Richard Ramaroson (ONERA, France)
Rosemary Renaut (Arizona State University, USA)
Reneé S. Renner (California State University at Chico, USA)
Paul Roe (Queensland University of Technology, Australia)
Alexey S. Rodionov (Russian Academy of Sciences, Russia)
Heather J. Ruskin (Dublin City University, Ireland)
Ole Saastad (Scali, Norway)
Muhammad Sarfraz (King Fahd University of Petroleum and Minerals, Saudi Arabia)
Edward Seidel (Louisiana State University, USA and Albert-Einstein-Institut, Potsdam, Germany)
Jie Shen (University of Michigan, USA)
Dale Shires (US Army Research Laboratory, USA)
Vaclav Skala (University of West Bohemia, Czech Republic)
Burton Smith (Cray, USA)
Masha Sosonkina (Ames Laboratory, USA)
Alexei Sourin (Nanyang Technological University, Singapore)
Elena Stankova (Institute for High Performance Computing and Data Bases, Russia)
Gunther Stuer (University of Antwerp, Belgium)
Kokichi Sugihara (University of Tokyo, Japan)
Boleslaw Szymanski (Rensselaer Polytechnic Institute, USA)
Ryszard Tadeusiewicz (AGH University of Science and Technology, Poland)
C.J. Kenneth Tan (OptimaNumerics, UK and Queen's University of Belfast, UK)
David Taniar (Monash University, Australia)
John Taylor (Streamline Computing, UK)
Ruppa K. Thulasiram (University of Manitoba, Canada)
Pavel Tvrdik (Czech Technical University, Czech Republic)
Putchong Uthayopas (Kasetsart University, Thailand)
Mario Valle (Swiss National Supercomputing Centre, Switzerland)
Marco Vanneschi (University of Pisa, Italy)
Piero Giorgio Verdini (University of Pisa and Istituto Nazionale di Fisica Nucleare, Italy)
Jesus Vigo-Aguiar (University of Salamanca, Spain)

Jens Volkert (University of Linz, Austria)
Koichi Wada (University of Tsukuba, Japan)
Stephen Wismath (University of Lethbridge, Canada)
Kevin Wadleigh (Hewlett Packard, USA)
Jerzy Wasniewski (Technical University of Denmark, Denmark)
Paul Watson (University of Newcastle Upon Tyne, UK)
Jan Weglarz (Poznan University of Technology, Poland)
Tim Wilkens (Advanced Micro Devices, USA)
Roman Wyrzykowski (Technical University of Czestochowa, Poland)
Jinchao Xu (Pennsylvania State University, USA)
Chee Yap (New York University, USA)
Osman Yasar (SUNY at Brockport, USA)
George Yee (National Research Council and Carleton University, Canada)
Yong Xue (Chinese Academy of Sciences, China)
Igor Zacharov (SGI Europe, Switzerland)
Xiaodong Zhang (College of William and Mary, USA)
Aledander Zhmakin (SoftImpact, Russia)
Krzysztof Zielinski (ICS UST / CYFRONET, Poland)
Albert Zomaya (University of Sydney, Australia)

Sponsoring Organizations

Institute of Electrical Engineers (IEE), UK
University of Perugia, Italy
University of Calgary, Canada
University of Minnesota, USA
Queen's University of Belfast, UK
The European Research Consortium for Informatics and Mathematics (ERCIM)
The 6th European Framework Project "Distributed European Infrastructure for Supercomputing Applications" (DEISA)
OptimaNumerics, UK
INTEL
AMD

Table of Contents – Part III

Workshop on Approaches or Methods of Security Engineering (AMSE 2006, Sess. A)

General Tracks

A First Approach to a Data Quality Model for Web Portals

Angelica Caro[1], Coral Calero[2], Ismael Caballero[2], and Mario Piattini[2]

[1] Universidad del Bio Bio, Departamento de Auditoria e Informática,
La Castilla s/n, Chillán, Chile
mcaro@ubiobio.cl

[2] ALARCOS Research Group,
Information Systems and Technologies Department,
UCLM-Soluziona Research and Development Institute,
University of Castilla-La Mancha, Paseo de la Universidad, 4 – 13071 Ciudad Real, Spain
{Coral.Calero, Ismael.Caballero, Mario.Piattini}@uclm.es

Abstract. The technological advances and the use of the internet have favoured the appearance of a great diversity of web applications, among them Web Portals. Through them, organizations develop their businesses in a really competitive environment. A decisive factor for this competitiveness is the assurance of data quality. In the last years, several research works on Web Data Quality have been developed. However, there is a lack of specific proposals for web portals data quality. Our aim is to develop a data quality model for web portals focused on three aspects: data quality expectations of data consumer, the software functionality of web portals and the web data quality attributes recompiled from a literature review. In this paper, we will present the first version of our model.

1 Introduction

In the last years, a growing interest in the subject of Data Quality (DQ) or Information Quality (IQ) has been generated because of the increase of interconnectivity of data producers and data consumers mainly due to the development of the internet and web technologies. The DQ/IQ is often defined as "fitness for use", i.e., the ability of a data collection to meet user requirements [1, 2]. Data Quality is a multi-dimensional concept [2], and in the DQ/IQ literature several frameworks providing categories and dimensions as a way of facing DQ/IQ problems can be found.

Research on DQ/IQ started in the context of information systems [1, 3] and it has been extended to contexts such as cooperative systems [4-6], data warehouses [7, 8] or electronic commerce [9, 10], among others.

Due to the characteristics of web applications and their differences from the traditional information systems, the community of researchers has recently started to deal with the subject of DQ/IQ on the web [11]. However, there are not works on DQ/IQ specifically developed for web portals. As the literature shows that DQ/IQ is very dependent on the context, we have centred our work on the definition of a Data Quality Model for web portals. To do so, we have used some works developed for differ-

M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 984 – 993, 2006.

ent contexts on the web but that can be partially applied or adapted to our particular context. For example, we have used the work of Yang et al., (2004) where a quality framework for web portals is proposed including data quality as a part of it.

As the concept of "fitness for use" is widely adopted in the literature (emphasizing the importance of taking into consideration the consumer viewpoint of quality), we have also considered, for the definition of our model, the data consumer viewpoint. First, we have combined the data quality expectations of data consumers with the software functionality of web portals. From the resultant matrix (data consumer expectations x functionalities), we have determined which web data quality attributes, recompiled in a literature review, can be applied.

The structure of this paper is as follows. In section 2, the components of our model are presented. In section 3, we will deeply describe the first version of our DQ/IQ Web Portal Model. Finally, in section 4 we will conclude with our general remarks and future work.

2 Model Components

Web Portals are emerging Internet-based applications that enable access to different sources (providers) through a single interface [12]. The primary objective of a portal software solution is to create a working environment where users can easily navigate in order to find the information they specifically need to perform their operational or strategic functions quickly as well as to make decisions [13], being responsibility of web portals' owners the achievement and maintenance of a high information quality state [14].

In this section, we will present the three basic aspects considerated to define our DQ/IQ model for web portals: the DQ/IQ attributes defined in the web context, the data consumer expectations about data quality, and web portals functionalities.

2.1 Data Consumer Expectations

When data management is conceptualized as a production process [1], we can identify three important roles in this process: (1) *data producers* (who generate data), (2) *data custodians* (who provide and manage resources for processing and storing data), and (3) *data consumers* (who access and use data for their tasks).

As in the context of web-based information systems, roles (1) and (2) can be developed by the same entity [11], for web portals context we identify two roles in the data management process: (1) *data producers-custodians*, and (2) *data consumers*.

So far, except for few works in DQ/IQ area, like [1, 2, 15, 16], most of the works on the subject have looked at quality from the data producer-custodian perspective. This perspective of quality differs from this in two important ways [15]:

- Data consumer has no control over the quality of available data.
- The aim of consumers is to find data that match their personal needs, rather than provide data that meet the needs of others.

Our proposal of a DQ/IQ model for web portals considers the data quality expectations of data consumer because, at the end, it is the consumer who will judge whether a data is fitted for use or not [16].

We will use the quality expectations of the data consumer on the Internet, proposed in [17]. These expectations are organized into six categories: Privacy, Content, Quality of values, Presentation, Improvement, and Commitment.

2.2 Web Portal Functionalities

A web portal is a system of data manufacturing where we can distinguish the two roles established in the previous subsection. Web portals present basic software functionalities to data consumer deploying their tasks and under our perspective, the consumer judges data by using the application functionalities. So, we used the web portal software functions that Collins proposes in [13] considering them as basics in our model. These functions are as follows: Data Points and Integration, Taxonomy, Search Capabilities, Help Features, Content Management, Process and Action, Collaboration and Communication, Personalization, Presentation, Administration, and Security. Behind these functions, the web portal encapsulates the producer-custodian role.

2.3 Web Data Quality Review

By using a DQ/IQ framework, organizations are able to define a model for data, to identify relevant quality attributes, to analyze attributes within both current and future contexts, to provide a guide to improve DQ/IQ and to solve data quality problems [18]. In the literature, we have found some proposals oriented to DQ/IQ on the web.

Among them, we can highlight those showed in table 1. Related to such proposals, we can conclude that there is no agreement concerning either the set of attributes or, in several cases, their meaning. This situation, probably, is a consequence of the different domains and author's focus of the studied works.

However, from this revision we captured several data quality attributes. The most considered are (we present between brackets different terms used for the same concept): Accuracy (Accurate), in 60% of the works; Completeness, in 50% of the works and Timeliness (Timely), in 40% of the works; Concise (Concise representation), Consistent (Consistent representation), Currency (Current), Interpretability, Relevance, Secure (Security), in 30% of the studies. Accessibility (Accessible), Amount of data (Appropriate amount of information), Availability, Credibility, Objectivity, Reputation, Source Reliability, Traceability (Traceable), Value added are stated in 20% of the works. Finally, Applicable, Clear, Comprehensive, Confidentiality, Content, Convenient, Correct, Customer Support, Degree of Duplicates, Degree of Granularity, Documentation, Understand ability (Ease of understanding), Expiration, Flexibility, Freshness, Importance, Information value, Maintainable, Novelty, Ontology, Pre-decision availability, Price, Reliability, Response time, Layout and design, Uniqueness, Validity, and Verifiability are only studied in 10 % of the works.

Summarizing the above-mentioned attributes, by means of similarity in their names and definitions, we have obtained a set of 28 attributes. Based on these DQ/IQ attributes we will try to identify which ones are applicable to the web portals context by classifying them into the matrix construed by the previous aspects (data consumer expectations x functionalities).

Table 1. Summary of web DQ/IQ framework in the literature

Author	Domain	Framework structure
[19]	Personal web sites	4 categories and 7 constructors
[20]	Data integration	3 classes and 22 of quality criterion
[10]	e-commerce	7 stages to modelling DQ problems
[9]	e-commerce	4 categories associated with 3 categories of data user requirements.
[21]	Web information systems (data evolution)	4 categories, 7 activities of DQ design and architecture to DQ management.
[6]	e-service cooperative	8 dimensions
[22]	Decision making	8 dimensions and 12 aspects related to (providers/consumers)
[23]	Web sites	4 dimensions and 16 attributes
[11]	DQ on the web	5 dimensions
[24]	Web sites	5 categories and 10 sub-categories
[25]	Organizational networks	6 stages to DQ analysis with several dimensions associated with each one
[26]	Data integration	2 factors and 4 metrics
[27]	Web information portals	2 dimensions

3 Relationships Between the Components of the Model

Based on the previous background, we will determine the relationship between the web portal functionalities and the quality expectations of data consumers. Then, we will present the definition of each function according to [13] and we will show their relationships (see figure 2).

- *Data Points and Integration.* They provide the ability to access information from a wide range of internal and external information sources and display the resulting information at the single point-of-access desktop. The expectations applied to this functionality are: *Content* (Consumers need a description of portal areas covered, use of published data, etc.), *Presentation* (formats, language, and others are very important for easy interpretation) and *Improvement* (users want to participate with their opinions in the portal improvements knowing the result of applying them).
- *Taxonomy.* It provides information context (including the organization-specific categories that reflect and support organization's business), we consider that the expectations of data consumer are: *Content* (consumers need a description of which data are published and how they should be used, easy-to-understand definitions of every important term, etc.), *Presentation* (formats and language in the taxonomy are very important for easy interpretation, users should expect to find instructions when reading the data), and *Improvement* (user should expect to convey his/her comments on data in the taxonomy and know the result of improvements).
- *Search Capabilities*. It provides several services for web portal users and needs searches across the enterprise, World Wide Web, and search engine catalogs and

indexes. The expectations applied to this functionality are: *Quality of values* (Data consumer should expect that the result of searches is correct, current and complete), *Presentation* (formats and language are important for consumers, for the search and for easy interpretation of results) and *Improvement* (consumer should expect to convey his/her comments on data in the taxonomy and know the result of improvements).

- *Help Features*. They provide help when using the web portal. The expectations applied to this functionality are: *Presentation* (formats, language, and others are very important for easy interpretation of help texts) and *Commitment* (consumer should be easily able to ask and obtain answer to any question regarding the proper use or meaning of data, update schedules, etc.).
- *Content Management*. This function supports content creation, authorization, and inclusion in (or exclusion from) web portal collections. The expectations applied to this functionality are: *Privacy* (it should exist privacy policy for all consumers to manage, to access sources and to guarantee web portals data), *Content* (consumers need a description of data collections, that all data needed for an intended use are provided, etc.), *Quality of values* (consumer should expect that all data values are correct, current and complete, unless otherwise stated), *Presentation* (formats and language should be appropriate for easy interpretation), *Improvement* (consumer should expect to convey his/her comments on contents and their management and know the result of the improvements) and *Commitment* (consumer should be easily able to ask and have any question regarding the proper use or meaning of data, update schedules, etc. answered).
- *Process and Action*. This function enables the web portal user to initiate and participate in a business process of portal owner. The expectations applied to this functionality are: *Privacy* (Data consumer should expect that there is a privacy policy to manage the data about the business on the portal), *Content* (Consumers should expect to find descriptions about the data published for the processes and actions, appropriate and inappropriate uses, that all data needed for the process and actions are provided, etc.), *Quality of values* (that all data associated to this function are correct, current and complete, unless otherwise stated), *Presentation* (formats, language, and others are very important for properly interpret data), *Improvement* (consumer should expect to convey his/her comments on contents and their management and know the result of improvements) and *Commitment* (consumer should be easily able to ask and to obtain answer to any questions regarding the proper use or meaning of data in a process or action, etc.).
- *Collaboration and Communication*. This function facilitates discussion, locating innovative ideas, and recognizing resourceful solutions. The expectations applied to this functionality are: *Privacy* (consumer should expect privacy policy for all consumers that participate in activities of this function), and *Commitment* (consumer should be easily able to ask and have any questions regarding the proper use or meaning of data for the collaboration and/or communication, etc).
- *Personalization*. This is a critical component to create a working environment that is organized and configured specifically to each user. The expectations applied to this functionality are: *Privacy* (consumer should expect privacy and security about their personalization data, profile, etc.), and *Quality of values* (data about user profile should be correct, current).
- *Presentation*. It provides both the knowledge desktop and the visual experience to the web portal user that encapsulates all of the portal's functionality. The expectations

applied to this functionality are: *Content* (the presentation of a web portal should include data about covered areas , appropriate and inappropriate uses, definitions, information about the sources, etc.), *Quality of values* (the data of this function should be correct, current and complete.), *Presentation* (formats, language, and others are very important for easy interpretation and appropriate use of portals data.) and *Improvement* (consumer should expect to convey his/her comments on contents and their management and know the result of the improvements).

- *Administration*. This function provides service for deploying maintenance activities or tasks associated with the web portal system. The expectations applied to this functionality are: *Privacy* (Data consumers need security for data about the portal administration) and *Quality of values* (Data about tasks or activities of administration should be correct and complete).
- *Security*. It provides a description of the levels of access that each user or groups of users are allowed for each portal application and software function included in the web portal. The expectations applied to this functionality are: *Privacy* (consumer need privacy policy about the data of the levels of access of data consumers.), *Quality of values* (data about the levels of access should be correct and current.) and *Presentation* (data about security should be in format and language for easy interpretation).

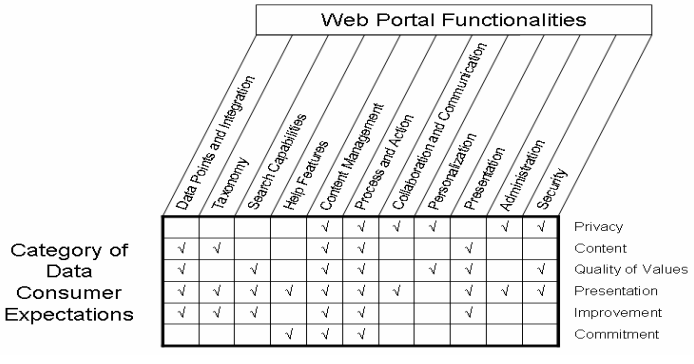

Fig. 2. Matrix stating the relationships between data consumer expectations and web portal functionalities

Concerning the relationships established in the matrix of figure 2, we can remark that *Presentation* is the category of data consumer expectation with more relations. This perfectly fits with the main goal of any web applications, which is to be useful and user-friendly for any kind of user.

The next step is to fill in each cell of the matrix with Web DQ/IQ attributes obtained from the study presented in 2.3. As a result of this, we have a subset of DQ/IQ attributes that can be used in a web portal to evaluate data quality. In table 2, we will show the most relevant attributes for each category of data consumer expectations.

To validate and complete this assignation we plan to work with portal data consumers through surveys and questionnaires. Once the validation is finished, we will reorganize the attributes obtaining the final version of the DQ/IQ web portal model.

Table 2. Web Data Quality attributes applied to web portal functionalities in each category

Category of Data Consumer Expectations	Web portal functionalities related to each category
Web DQ/IQ attributes applying almost one functionality in each category	
Privacy	Content management, Process and actions, Collaboration and Communication, Personalization, Administration, Security
Security	
Content	Data Points and Integration, Taxonomy, Content management, Process and actions, Presentation
Accessibility, Currency, Amount of data, Understandability, Relevance, Concise Representation, Validity, Traceability, Completeness, Reliability, Credibility, Timeliness, Availability, Documentation, Specialization, Interpretability, Easy to use	
Quality of data	Data Points and Integration, Search Capabilities, Content management, Process and actions, Personalization, Presentation, Security
Accessibility, Currency, Amount of data, Credibility, Understandability, Accuracy, Expiration, Novelty, Relevance, Validity, Concise Representation, Completeness, Reliability, Availability, Documentation, Duplicity, Specialization, Interpretability, Objectivity, Relevance, Reputation, Traceability, Utility, Value-added, Easy to use	
Presentation	Data Points and Integration, Taxonomy, Search Capabilities, Help Features, Content management, Process and actions, Collaboration and Communication, Presentation, Administration, Security
Amount of data, Completeness, Understandability, Easy to use, Concise Representation, Consistent Representation, Validity, Relevance, Interpretability, User support, Availability, Specialization, Flexibility	
Improvement	Data Points and Integration, Taxonomy, Search Capabilities, Content management, Process and actions, Presentation
Accessibility, Reliability, Credibility, Understandability, User support, Traceability	
Commitment	Help Features, Content management, Process and actions
Accessibility, Reliability, User support,	

4 Validation of the Model

In order to valid our model we plan to elaborate a survey to check the DQ/IQ attributes identified as relevant to the web portals. We will use the *Principles of Survey Research* proposed in [28] where is said that a survey is part of a larger process and recognize that it is not just the instrument for gathering information. In this work the authors identify ten activities in the survey process.

At this moment we are developing the first activities in our survey process. In particular *setting specific and measurable objectives* (in our case this phase consists in

checking the DQ/IQ attributes identified as relevant to the web portals and in obtaining other than were not considered), *planning and scheduling the survey, ensuring that appropriate resources are available* and *designing the data collection instrument.*

As survey design we have selected the descriptive design because we try to describe a phenomenon of interest [29] (in our case is to describe the DQ/IQ attributes more relevant for web portal data consumers). We plan to made a questionnaire for each one of the web portal functionalities presented previously. As it is quite impossible to survey the entire population [29], we are developing a web application to be linked in a web portal (www.castillalamancha.es). In that way, the users connected to this portal will be invited to answer some questions (selected randomly between the eleven questionnaires). So, each questionnaire will be constructed for each subject of the survey with the aim of having a correct distribution in the amount of answers given to each question.

The application will have three modules: an administrator module (through it the researcher can generate the questionnaires deciding the number of questions, the type of answer, etc.), an analyzer module (that shows the results: statistics, graphics, ranking of responses, etc.), and a gather module (that presents the questions to the users). So, we will ask each subject about general demographic questions (as the expertise in the use of portals, expertise in technologies, range of age, sex, etc.) together with thirty questions selected randomly from all questions in the eleven questionnaires. When we have enough responses for each question in our questionnaires we will analyze the responses for obtaining a minimum and necessary set of DQ/IQ attributes for each aspect of our model. This set of attributes will be used in order to elaborate a complete framework for evaluating the DQ/IQ of a web portal. For example, we plan to give the minimum value necessary for each attribute for assuring the DQ/IQ quality. If this value is not achieved for some of the attributes, the framework will give some corrective actions applicable in order to have the correct level of quality.

5 Conclusions

The great majority of works found in the literature show that data quality or information quality is very dependent on the context. The increase of the interest in the development of web applications has implied either the appearance of new proposals of frameworks, methodologies and evaluation methods of DQ/IQ or the adaptation of the already-existing ones from other contexts. However, in the web portal context, data quality frameworks do not exist.

In this paper, we have presented a proposal that combines three aspects: (1) a set of web data quality attributes resulting from a data quality literature survey that can be applicable and useful for a web portal, (2) the data quality expectations of data consumer on the Internet, and (3) the basic functionalities for a web portal. These aspects have been related by obtaining a first set of data quality attributes for the different data consumer expectations X functionalities.

Our future work, now in progress, consists of validating and refining this model. First of all, it is necessary to check these DQ/IQ attributes with data consumers in a web portal, for this we plan to make a survey as was presented in the previous section.

Then, once we have validated the model, we will define a framework including the necessary elements to evaluate a DQ/IQ in a web portal. Our aim is to obtain a flexible framework where the data consumer can select the attributes used to evaluate the quality of data in a web portal, depending on the existing functionalities and their personal data quality expectations.

Acknowledgements

This research is part of the following projects: CALIPO (TIC2003-07804-C05-03) supported by the Dirección General de Investigación of the Ministerio de Ciencia y Tecnología (Spain) and DIMENSIONS (PBC-05-012-1) supported by FEDER and by the "Consejería de Educación y Ciencia, Junta de Comunidades de Castilla-La Mancha" (Spain) and CALIPSO (TIN2005-24055-E).

References

1. Strong, D., Y. Lee, and R. Wang, *Data Quality in Context.* Communications of the ACM, 1997. Vol. 40, Nº 5: p. 103 -110.
2. Cappiello, C., C. Francalanci, and B. Pernici. *Data quality assessment from the user´s perspective*. in *International Workshop on Information Quality in Information Systems, (IQIS2004)*. 2004. Paris, Francia: ACM.
3. Lee, Y., *AIMQ: a methodology for information quality assessment.* Information and Management. Elsevier Science, 2002: p. 133-146.
4. Winkler, W., *Methods for evaluating and creating data quality.* Information Systems, 2004. Nº 29: p. 531-550.
5. Marchetti, C., et al. *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*. 2003.
6. Fugini, M., et al., *Data Quality in Cooperative Web Information Systems.* 2002.
7. Zhu, Y. 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*. 2002.
8. Bouzeghoub, M. and Z. Kedad, *Quality in Data Warehousing*, in *Information and Database Quality*, M. Piattini, C. Calero, and M. Genero, Editors. 2001, Kluwer Academic Publishers.
9. Katerattanakul, P. and K. Siau, *Information quality in internet commerce desing*, in *Information and Database Quality*, M. Piattini, C. Calero, and M. Genero, Editors. 2001, Kluwer Academic Publishers.
10. Aboelmeged, M. *A Soft System Perspective on Information Quality in Electronic Commerce*. in *Proceeding of the Fifth Conference on Information Quality*. 2000.
11. Gertz, M., et al., *Report on the Dagstuhl Seminar "Data Quality on the Web".* SIGMOD Record, 2004. vol. 33, Nº 1: p. 127-132.
12. Mahdavi, M., J. Shepherd, and B. Benatallah. *A Collaborative Approach for Caching Dynamic Data in Portal Applications*. in *Proceedings of the fifteenth conference on Australian database*. 2004.
13. Collins, H., *Corporate Portal Definition and Features*. 2001: AMACOM.

14. Kopcso, D., L. Pipino, and W. Rybolt. *The Assesment of Web Site Quality*. in *Proceeding of the Fifth International Conference on Information Quality*. 2000.
15. Burgess, M., N. Fiddian, and W. Gray. *Quality Measures and The Information Consumer*. in *IQ2004*. 2004.
16. Wang, R. and D. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumer.* Journal of Management Information Systems, 1996. 12(4): p. 5-33.
17. Redman, T., *Data Quality: The Field Guide*. 2001: Digital Press.
18. Kerr, K. and T. Norris. *The Development of a Healthcare Data Quality Framework and Strategy*. in *IQ2004*. 2004.
19. Katerattanakul, P. and K. Siau. *Measuring Information Quality of Web Sites: Development of an Instrument*. in *Proceeding of the 20th International Conference on Information System*. 1999.
20. Naumann, F. and C. Rolker. *Assesment Methods for Information Quality Criteria*. in *Proceeding of the Fifth International Conference on Information Quality*. 2000.
21. Pernici, B. and M. Scannapieco. *Data Quality in Web Information Systems*. in *Proceeding of the 21st International Conference on Conceptual Modeling*. 2002.
22. Graefe, G. *Incredible Information on the Internet: Biased Information Provision and a Lack of Credibility as a Cause of Insufficient Information Quality*. in *Proceeding of the Eighth International Conference on Information Quality*. 2003.
23. Eppler, M., 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 the Eighth International Conference on Information Quality*. 2003.
24. Moustakis, V., et al. *Website Quality Assesment Criteria*. in *Proceeding of the Ninth International Conference on Information Quality*. 2004.
25. Melkas, H. *Analyzing Information Quality in Virtual service Networks with Qualitative Interview Data*. in *Proceeding of the Ninth International Conference on Information Quality*. 2004.
26. Bouzeghoub, M. and V. Peralta. *A Framework for Analysis of data Freshness*. in, *IQIS2004*. 2004. Paris, France: ACM.
27. Yang, Z., et al., *Development and validation of an instrument to measure user perceived service quality of information presenting Web portals.* Information and Ma-nagement. Elsevier Science, 2004. 42: p. 575-589.
28. Pfleeger, S. and B. Kitchenham, *Principles of Survey Research Part1: Turning Le-mons into Lemonade.* Software Engineering Notes, 2001. 26(6): p. 16-18.
29. Pfleeger, S. and B. Kitchenham, *Principles of Survey Research Part2: Designing a Survey.* Software Engineering Notes, 2002. 27(1): p. 18-20.